

WHAT MORAL RESPONSIBILITY REQUIRES

BY

WILLIAM SIMKULET

Submitted to the graduate degree program in Philosophy in the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Chairperson Dale Dorsey, Ph. D.

John Bricke, Ph. D.

Thomas Tuozzo, Ph. D.

Ben Eggleston, Ph. D.

James Hartman, Ph. D.

Date Defended: March 11 2011

The Dissertation Committee for William Simkulet certifies that this is the approved version of the following dissertation:

WHAT MORAL RESPONSIBILITY REQUIRES

Chairperson Dale Dorsey, Ph. D.

Date approved: March 11 2011

Abstract: What Moral Responsibility Requires

The primary goal of this dissertation is to articulate and defend a robust commonsense libertarian theory of moral responsibility; that moral agents are the causes, and owners, of their actions, and in virtue of this it is appropriate to hold them praiseworthy or blameworthy for what they do.

Here, I critique and defend two commonsense principles concerning moral responsibility – the control principle, and the principle of alternate possibilities. In recent years these principles have come under attack from philosophers seeking to propose a theory of moral responsibility consistent with a deterministic worldview. The existence of moral luck would mean that the control principle is false; I argue that moral luck is impossible. Harry Frankfurt famously presents a supposed counter-example to the principle of alternate possibilities; I argue Frankfurt's case turns on an equivocation between alternate possibilities and alternate outcomes.

I contend that moral responsibility requires an agent to have full control of her actions, to be the author of what she is praiseworthy or blameworthy for. This view requires indeterminism of a special kind, agent-causation (or something very much like it), where something is an agent-cause if and only if at a given time, it, and only it, determines its actions, and was not determined to act in this way by outside forces. Only agent-causes can be truly responsible for their actions because only the actions of agent-causes can be traced back to them and no further. And finally I argue we have good reason to believe we are such agent-causes.

TABLE OF CONTENTS:

Title Page.....	page i
Acceptance Page.....	page ii
Abstract.....	page iii
Table of Contents.....	page iv
Dissertation	
Chapter 1: Introduction.....	page 1
Chapter 2: The Control Principle.....	page 9
Chapter 3: The Principle of Alternate Possibilities.....	page 44
Chapter 4: What Moral Responsibility Requires.....	page 70
Chapter 5: Conclusion.....	page 105
Bibliography.....	page 115

William Simkulet

Dissertation: *What Moral Responsibility Requires*

Chapter 1: Introduction

1.1 Prospectus

The primary goal of this work is to defend a robust, commonsense notion of moral responsibility from critics whose work has been taken to weaken our confidence in our moral beliefs on the matter. I contend that the commonsense moral beliefs that most of us come to the table with, so to speak, are maximally consistent with a *libertarian* theory of moral responsibility, that moral agents are morally responsible for their actions because they are the causes, and thus owners, of those actions.

In this dissertation, I discuss three kinds of arguments that, if successful, would each give us good reason to revise a substantial portion of our commonsense beliefs, including the vast majority of our moral beliefs, and reject the vast majority of our moral intuitions. In each case, I argue that these criticisms fail to shake these beliefs, beliefs which are best understood as the foundation for a libertarian theory of moral responsibility and moral agency.

I begin in chapter 2 with a defense of the *control principle*, according to which moral responsibility requires control over one's actions; control of the kind necessary for libertarian ownership. According to the problem of moral luck, all of our actions are a matter of luck, and thus none of us has the control required for moral responsibility. I argue that moral luck is impossible, and that although moral agents lack control over many aspects of their life, they have control over what matters – their free choices.

Chapter 3 is a defense of the *principle of alternate possibilities*, which states that moral agents can only be held morally responsible if they could have done otherwise. Harry Frankfurt,

in his now famous article “Alternate Possibilities and Moral Responsibility”, argues that the principle of alternate possibilities is false, and constructs a case that purports to show an agent who is undeniably morally responsible, despite lacking alternate possibilities.¹ Frankfurt’s infamous case fails in its task, and I argue that more recent Frankfurt-style cases also fail.

In chapter 4, I argue that *indeterminism* is logically possible, and that we have good reason to believe that it is true of the actual world. *Determinism* is the theory that the actual state of affairs, coupled with the actual laws of nature, determine a unique future, and that there are no other possible futures given the actual state of affairs and the actual laws of nature. In contrast, *indeterminism* is the theory that at any given time, the actual state of affairs and the actual laws of nature may cause one of several possible futures to become actual. If the control principle and the principle of alternate possibilities are true, moral responsibility requires an indeterministic world. To the extent that we believe moral agents exist, and that some people are actually morally responsible for their actions, we are committed to the view that indeterminism is true of the actual world.

In the final chapter of this dissertation, I propose several avenues of investigation for both the critic and the supporter of libertarianism. For the critic, I offer new approaches that, if successful, would give us reason to reject our foundational moral beliefs with which libertarianism is so consistent. Although I am doubtful these approaches will lead to success, I believe they are the most likely approaches to lead to success. In contrast, for the libertarian I sketch the framework for a robust theory of moral responsibility, and a method for calculating how praiseworthy or blameworthy moral agents are in virtue of what they have complete control over, their free choices.

¹ Frankfurt, Harry G., 1969, “Alternate Possibilities and Moral Responsibility”, *The Journal of Philosophy*, Vol. 66, No. 23: 829-839.

1.2 Method

The primary goal of this work is a meta-ethical inquiry into the concept of moral responsibility, with a special focus on what conditions are necessary for an agent to be morally responsible. The main body of this dissertation is concerned with sketching a consistent interpretation of this concept and its application. A secondary goal for this work is to show that we have good reason to believe that moral responsibility is possible at the actual world.

Because this primary goal of this dissertation is an inquiry into the actual concept of moral responsibility people have, I approach the subject in a manner similar to Donald Davidson's radical interpretation, taking the stance that the best way to interpret our beliefs and ideas about moral responsibility is to assume (a) that our beliefs are coherent and generally internally consistent with one another, and (b) that they correspond to the actual world in some non-arbitrary, non-coincidental way. The best interpretation of our concept of moral responsibility, then, is like the best interpretation of any other concept or belief – the one that makes us out to be right about most things (by weight, not number²).

It is, of course, possible that there is no coherent concept of moral responsibility, that different people have different concepts of moral responsibility, or that concept of moral responsibility that the majority of people have in mind is impossible at the actual world such that all of our ascriptions of moral praise and blame are false.

Despite this fact, to avoid crumbling into an abyss of radical skepticism, we must assume that people are right about most things *that matter*. While we can tell an evolutionary story to

² In order to avoid radical skepticism, and to engage in philosophical debate at all, we must make certain assumptions about the world, most notably that we are right about most things. However, this assumption that any given belief is more likely to be true than false, or that were we to count all of our beliefs then the majority of them would be true. Rather, it is to assume that the beliefs that we are most confident in, and the beliefs that we regularly use to help us in everyday life are most likely correct. We can, of course, consider whether or not certain portions of these beliefs are correct, but to seriously entertain the notion that a significant majority of these beliefs are false is to neuter one's ability to act as a rational agent.

this effect (roughly, people with numerous false beliefs about the matters that concern themselves on a regular basis are less likely to survive than those with largely true beliefs about those matters), to even begin to properly interpret others we must assume their beliefs are coherent and correspond to the actual world. Moral philosophy is, by definition, concerned with what we *ought* do, and to the extent that our moral beliefs guide our *every action*, we have good reason to believe they are both consistent and represent facts about the world. Thus, our moral beliefs are at least comparable to, if not more basic than, our beliefs about the physical world.

In virtue of this, I believe we have good reason to think that if something must be true of the physical world for the bulk of our moral beliefs to be true, then that thing *is* true of the physical world. In a sense, our moral beliefs can be taken as *ad hoc* indicators of physical truths about the world insofar as our moral beliefs are assumed true, and can be, such as in the case of our beliefs about moral responsibility, dependent upon certain truths about the physical world.

1.3 Terminology

In this section, I will define some of the important terms that will be used in this dissertation. Defined words are italicized. *Ethics* is the branch of philosophy concerned with answering the question “What ought I do?”. The act of asking the question implies two things. First, that “ought” (at least sometimes) implies “can”; the act of asking the question is pointless unless one can do as one ought. Second, the question also implies that one can do otherwise, that one can fail to act as they should. If moral success was guaranteed, and one couldn’t help but do what one should, then the question is pointless. These implications suggest that if moral facts exist, then we know certain facts about the world – we know that persons can, at least some times, do the things that they should, and they can also, at least some times), fail to do what they should.

To be *morally responsible* for something is to stand in some relationship to that thing, such that one is the appropriate object of moral praise or blame in virtue of that relationship. If the thing is morally good, the agent is praiseworthy, if it is morally bad, the agent is blameworthy. Galen Strawson calls this “true moral responsibility”³, and contends that for one to be truly morally responsible for something is for it to “*makes sense*” for one to be rewarded in heaven or punished in hell for that something.⁴ Only moral agents can be morally responsible; only moral agents can be morally praiseworthy or blameworthy.

To be morally *praiseworthy* for something is to be the appropriate object of praise in virtue of one’s moral responsibility for that thing, all things being equal. To be morally *blameworthy* for something is to be the appropriate object of blame in virtue of one’s moral responsibility for that thing, all things being equal.⁵

Moral responsibility can be contrasted with various kinds of non-moral responsibility. For example, moral and non-moral agents alike can be *causally responsible* for something, where to be causally responsible is to play a causal role in bringing about that something. To be *legally responsible* for something is to be such that a government can reward or punish you for that something under their laws.

Being morally responsible is also different from *taking responsibility* for something. The term “taking responsibility” has two conventional and somewhat misleading usages – in one sense, to take responsibility for something is to acknowledge one’s actual moral responsibility for that thing and to take steps to accept blame or praise in virtue of it. For example, one might

³ See Strawson, Galen 1994.

⁴ As Strawson puts it “The stress on the words ‘makes sense’ is important, for one certainly does not have to believe in any version of the story of heaven and hell in order to understand the notion of true moral responsibility that it is being used to illustrate.”

⁵ To be praiseworthy or blameworthy *is* not to be such that it is appropriate under all circumstances to be praised or blamed. There are circumstances when one ought not openly act to blame or punish the blameworthy, or praise or reward the praiseworthy. It is to be the appropriate object of praise or blame respectively, all things being equal.

admit to accidentally breaking something due to one's carelessness and offer to pay for it. The second sense in which one can take responsibility for something is to treat one's self as if one is morally responsible for that thing *regardless* of one's actual moral responsibility. For example, one might take responsibility for a dead relative's children and raise them despite no clear moral obligation to do so. One can also take responsibility in this way for something that it is *wholly* inappropriate to treat one as morally responsible for. For example, one might take responsibility for something she didn't do for some utilitarian gain, or to protect someone else.⁶

Just as things can be responsible in ways other than morally, things can be praiseworthy or blameworthy in non-moral ways as well. For example, a car might be praiseworthy to the extent that it is instrumentally valuable in achieving some goal, and blameworthy insofar as it fails to contribute to some other goal. It might be appropriate to praise a single car insofar as it requires little regular maintenance, is relatively safe for passengers, and/or is suitable for off-road driving, while at the same time it may be appropriate to blame the car for being fuel inefficient, for high carbon emissions, and/or for being difficult to get parts for. The difference between being morally praiseworthy or blameworthy and being praiseworthy or blameworthy in

⁶ In the article 1998 "Morally Responsible People without Freedom," John Martin Fischer and Mark Ravizza contend that agents become morally responsible by taking responsibility. For Fischer and Ravizza, agents are morally responsible for their behavior if and only if they own the mechanisms that cause their behavior, and they come to own these mechanisms by *taking responsibility* for them, where taking responsibility for a mechanism doesn't require knowledge or acceptance of the mechanism itself, but only embracing responsibility for the kinds of behavior the mechanism causes. Taking responsibility in this way doesn't require a direct or consistent admission of guilt for this behavior, rather it only requires that the agent accepts herself as an appropriate object of certain kinds of reactive attitudes because of the behavior of the mechanism.

Fischer and Ravizza's account of how one comes to be morally responsible is at odds with our commonsense beliefs – first, it seems to postulate an erroneous means of coming to own something – for Fischer and Ravizza, one owns something simply by treating oneself as if one owned it. It is, at the very least, fraud, if not downright theft. Their account of property acquisition is easily contrasted with John Locke's account of ownership. For Locke, one owns one's labor because one is the creator, cause, or author, of one's labor, while for Fischer and Ravizza one comes to take responsibility either because (a) one finds it practical to be treated as the owner of one's actions, or (b) one has been subject to (apparently erroneous) "moral education" that has taught one that one is the author, and owner, of one's actions. Fischer and Ravizza's approach is, in part, motivated by their belief that moral responsibility is compatible with determinism, and traditional accounts of ownership, such as the kind Locke discusses, are incompatible with determinism. Unfortunately, their account of "moral responsibility" is substantially at odds with commonsense notions of ownership and moral responsibility.

light of these other standards is that the standard moral agents are judged morally praiseworthy or blameworthy for is an intrinsically valuable one, while these other standards are merely instrumentally valuable, depending upon one's goals and the circumstances one finds oneself in. This is to say that it is *always* good to be morally praiseworthy, and *always* bad to be morally blameworthy, but it is *not* always important for a car to be able to drive off-road, and not always bad for it to be fuel inefficient (say, if there is a fuel surplus, etc.).

Libertarianism is the theory that moral agents are the authors, or sole causes, of their actions, and in virtue of this fact, *own* their actions, and can be morally responsible for them. According to this view, the morally relevant relationship moral responsibility is concerned with is a direct-ownership relationship – moral agents *own* their actions because they are the causes of their actions. Libertarianism is an *incompatibilist* theory because it holds that what an agent does is up to her, and not causally determined by antecedent circumstances alone. *Incompatibilism* is the theory that moral responsibility is incompatible with *determinism*, the theory that the actual past, coupled with the actual laws of nature, necessitates one possible future. According to the libertarian, moral responsibility requires *indeterminism* to be true at the actual world, where indeterminism is the theory that their actual state of affairs, coupled with the actual laws of nature, could produce one of multiple possible futures.

Compatibilism is the theory that moral responsibility is consistent with determinism. *Compatibilists* maintain that moral responsibility is possible if determinism is true. *Hard compatibilism* is the theory that moral responsibility *requires* determinism. *Hard incompatibilism* (sometimes referred to as *hard determinism*) is the theory that moral responsibility requires indeterminism, that determinism is true, and thus no one is actually morally responsible for anything.

1.4 Next

I contend that libertarianism is maximally consistent with our commonsense moral beliefs and intuitions. In the next two chapters, I look at philosophical arguments meant to show that our beliefs are at odds with libertarianism. In chapter 2, I argue that although we admit human beings lack complete control over every aspect of their lives, we believe they have complete control over their actions, and that in virtue of this it makes sense to say that they are morally responsible for those actions. In chapter 3, I defend the position that our moral beliefs and intuitions indicate that moral responsibility requires alternate possibilities; that is to say that moral responsibility is only possible in a world where indeterminism is true. In contrast, in chapter 4 I argue that the indeterminism our moral intuitions requires is not incoherent, that it is theoretically possible, and in virtue of the assumption necessary to interpret our concept of moral responsibility, we have good reason to think that the indeterminism required for moral responsibility occurs regularly in the actual world.

Chapter 2: The Control Principle

2.1 Introduction

To be morally responsible for something is to stand in a relationship to that thing such that, in virtue of that relationship, one can be either praiseworthy or blameworthy depending on whether that thing is good or bad. According to our commonsense moral beliefs it doesn't make sense to hold someone blameworthy or praiseworthy for something unless that thing was *up to that person*, and not merely a matter of luck or chance. In other words, it wouldn't make sense to blame someone for something that they had no control over.

For example, given what we believe about time, it is impossible for someone to be either blameworthy or praiseworthy for something that happened before they were born. Similarly, it is impossible for someone to be morally responsible for their own birth. Furthermore, intuitively, if some event prior to a person's being born causally determined them to act in a certain manner, that person is not morally responsible for what they are caused to do in this way.

This idea can be captured by the following principle:

Control Principle: Moral responsibility requires control. For one to be morally responsible for some thing, one must be in control of that thing. (CP from now on.)

According to libertarianism, moral responsibility requires a kind of ownership; moral agents are morally responsible for their actions because they *own* their actions, and they own their actions because they are the unique creators of their actions; when a moral agent acts, her actions are *up to her* and her alone. John Locke famously argued that one can come to own things in the world by mixing out labor with them. However, for Locke, this ownership is derived from your prior ownership of your labor, which you own because you are the creator of your own labor.

Sometimes it makes sense to say that moral agents are morally responsible for the consequences of their actions, but this is derivative of, and contingent on, said their ownership of their actions, much as how, for Locke, one's ownership of a concrete physical object, such as an apple you pick with your labor, is derived from the mixing of one's labor with said objects.

Libertarianism is maximally consistent with the control principle. Until recently, most moral philosophers accepted that some form of the control principle was true. However, the *problem of moral luck* threatens to undermine the plausibility of the control principle, and with it the plausibility of libertarianism. The problem can be summed up as follows: (1) according to the control principle, moral responsibility requires control and thus cannot be a matter of luck, (2) everything is a matter of luck, thus (3) moral responsibility is impossible. The problem arises because (1) and (2) are *prima facie* true, but the conclusion, (3), is *prima facie* false.

In this chapter I argue that the kind of control required for moral responsibility is not inconsistent with the way in which everything can be said to be a matter of luck. Although it may be a matter of luck that any given *set* of affairs exist at a given time, given this set of affairs, a moral agent may still have control over a subset of following sets of affairs.

2.2 Aristotle and the Control Principle

Aristotle is often credited with presenting the first theory of moral responsibility.⁷ For Aristotle, to be morally responsible for something is to be the appropriate object of praise or blame if that something warrants it. Early in book 3 of the *Nicomachean Ethics* Aristotle claims only an agent whose action is voluntary is an appropriate object of praise and blame.⁸ For

⁷ Randall Curren and Jean Roberts have challenged the view that the praiseworthiness and blameworthiness discussed in the *Nicomachean Ethics* is anything like our contemporary account of moral responsibility. If true, then Aristotle's account may be at odds with our contemporary normative beliefs and intuitions. However, Susan Meyer offers a stalwart defense of the traditional view in the early chapters of her *Aristotle on Moral Responsibility*, so we have good reason to think our contemporary interpretation mirrors Aristotle's own view.

⁸ Aristotle, *Nicomachean Ethics*, 1109b30-1110a4

something to be voluntary, he says, it must satisfy two conditions – a control condition and an awareness condition.⁹

Aristotle's control condition is first and foremost an origin condition – for an agent's action to be voluntary, it must originate in the agent – it cannot be compelled by factors outside of her control. If something external to the agent completely causally determines her action, for Aristotle, it doesn't come from the agent herself, and thus she isn't morally responsible for what she is caused to do. Strictly speaking, her actions are not her own in such a case. Similarly, if one's actions were completely causally determined by the sort of thing she is, they would be equally outside of her control and merely a byproduct of external forces – whatever it is that made her what she is. Furthermore, if one's actions are caused by chance or luck, and not the agent herself, they are equally outside of her control. To satisfy the control condition, one's actions need to be up to the agent alone, such that she is neither causally determined to do them, nor that it is merely blind, arbitrary luck that she acts in the way she does.

To satisfy the control condition, one's actions cannot be completely causally determined by the past, nor can they be the result of luck over which one has no control, such as quantum indeterminacy. Many philosophers worry that these conditions are mutually exclusive, but despite this it is at least *prima facie* intuitively plausible, for example, that agents can act *for reasons* without being causally determined to act for those reasons. Similarly, it is *prima facie* intuitively plausible that agents can choose to act for no reason at all, and so long as the choice itself wasn't the result of luck, said agents can be, and often are, blameworthy for such actions.

To satisfy the control condition is to have control over *whether* you act, however this alone is insufficient for moral responsibility as without some substantive rationale for choosing one action over another, acting would be arbitrary. Imagine an agent who is put in a position

⁹ Aristotle, *Nicomachean Ethics*, 1110a-1111b4

where she has the option of acting or not acting, but she lacks any beliefs about what it would be to act or not act, and what either would produce in the actual world. Such an agent has no substantive rationale for choosing to act or not act, and thus her choice is inherently arbitrary. At first glance, we might say that such an agent is morally responsible for the consequences of what she chose to do, but this intuition is based on the fact that we generally assume that people know what they are doing. By assumption, to that agent, acting and not acting are both unknowns and she is forced to do one of them. The problem with holding such an agent morally responsible for what she does is that in an important way she lacks control over what it is she causes because she is ignorant of what acting or not acting actually is. Not only does she lack any beliefs that suggest that acting would cause a better or worse scenario than not acting, but from her position acting and not acting are indistinguishable. To hold her morally responsible for her choice to act or not act in such a case would be to hold her responsible for something that is inherently arbitrary, as she has no way to differentiate between her two options, but must do one of them.

The idea that moral responsibility cannot be determined arbitrarily is, in many ways, even more foundational than the control principle, and can be captured by the following principle:

The non-arbitrary principle: Moral responsibility is not determined arbitrarily. Whether one is morally responsible for something is not a matter of chance. (NP from now on.)

Aristotle's second condition – the awareness condition – helps bridge the gap between the control condition and NP. To satisfy the awareness condition, the agent must be nontrivially aware of what she does in order to be morally responsible for it. But as with the control condition, merely satisfying the awareness condition is not sufficient for moral responsibility. If we are but puppets of external forces and luck, the puppet who is aware of what the puppeteer

makes her do is no less a slave, no less in control, and it doesn't make sense to attribute her actions to her rather than the puppeteer.

If an agent satisfies both the control condition and the awareness condition, she is in control of both *whether* she acts and *how* she acts. If an agent has this kind of control over what she does, then she can not only choose to act or not act, but she can choose to do so on the basis of what she knows she would be doing or avoiding.¹⁰ This is to say that she can act *for reasons*, and thus non-arbitrarily. It is *this* kind of control that is *prima facie* necessary for moral responsibility for Aristotle, and this *just is* the kind of control the control principle requires for moral responsibility.

2.3 Moral Luck

The control principle's *prima facie* intuitive plausibility comes, in large part, from its assertion that our moral responsibility, and what we deserve, morally speaking, is not arbitrary or random; rather it's completely *up to us* whether or not we are blameworthy or praiseworthy (independent of whether others actually blame or praise us).

Despite this, it seems as if we are perfectly content to say that people are morally responsible for things that are outside of their control – a matter of luck. For example, it is generally accepted that moral agents are morally responsible for the consequences of their actions, despite the fact that the actual consequences of one's actions are beyond one's control.

Moral luck would occur if someone is morally responsible for something that is outside of their control. The control principle precludes the possibility of moral luck; thus if moral luck is possible, CP is false. The *problem of moral luck* is that we seem to hold two contradictory

¹⁰ On this model it is possible for an agent to act on the basis of what she knows. However, it seems equally possible that an agent can choose to act (a) in absence of knowledge about what she is doing, or (b) in spite of what she knows. In these cases, what makes such an agent morally responsible is that although she may not know, or care, what she is doing, she *is aware* that she doesn't know, or care what she is doing. Her actions are still voluntary.

positions – first, that moral responsibility requires control, and second that moral agents are occasionally morally responsible for things that are outside of their control, such as their character traits, the results of their actions, etc.

For the rest of this chapter, I look at the problem of moral luck. I begin by looking at the problem of moral luck, and distinguishing between *moral luck* and what I call *record luck*, or luck in regards to one's moral record. I argue record luck is *not* a kind of moral luck, and thus the existence of record luck doesn't show that the control principle is false. Next I discuss the four kinds of moral luck that some have said exist – *resultant luck*, *circumstantial luck*, *constitutive luck*, and *causal luck* – and argue that the luck involved in the types of cases moral luck advocates discuss does not contribute to our actual blameworthiness or praiseworthiness, although it may contribute to our *apparent* blameworthiness or praiseworthiness. This is to say that it may be a matter of luck whether or not one is praised for doing good, or blamed for doing bad, but that it is *not* a matter of luck whether one is praiseworthy or blameworthy.

2.3.1 The Problem of Moral Luck

The problem of moral luck is that we tend to be committed to two contradictory propositions – (1) the control principle is true – moral agents are only morally responsible for what is in their control, and (2) people can be morally responsible for things that are, strictly speaking, outside of their control, such as their character, the results of their actions, or their being put into “no win” situations where, up to that point, they have done nothing wrong, but after that point, whatever choice they make, they will be morally blameworthy for something.

In *Luck and Moral Responsibility*, Michael Zimmerman frames the problem of moral luck as arising from the following argument against the possibility of morality:

1. A person P is morally responsible for an event *e*'s occurring only if *e*'s occurring was not a matter of luck.

2. No event is such that its occurring is not a matter of luck.

Therefore

3. No event is such that P is morally responsible for its occurring. (1987, 374)

The problem, Zimmerman explains, is that the argument seems valid, and both premises seem true, but the conclusion is *prima facie* false. The first premise follows from the control principle, while the second follows largely from our at best limited control over the world, control that itself is a matter of luck insofar as it rests upon the contingency of our birth. In light of the validity of the argument, philosophers have three possibilities – (1) they can accept the conclusion, that no one is morally responsible for anything¹¹, (2) they can accept the possibility of moral luck and accept the falsity of the control principle¹², or (3) they can reject the possibility of moral luck and show that there is at least one situation where an agent's moral responsibility is in no way a byproduct of luck. I take approach (3) below.

2.3.1.1 Terminology

Before we begin, I wish to settle on some terminology to frame the upcoming discussion. Let's begin with *moral responsibility*. For Aristotle, to be *morally responsible* is to be the appropriate object of praise or blame for that thing, if that thing warrants it. Derk Pereboom adds to this concept of moral responsibility by pointing out that it makes sense to say that an agent is morally responsible even when she is not praiseworthy or blameworthy, such as “when

¹¹ Zimmerman contends that Joel Feinberg does this in his 1970 article *Doing and Deserving*, pg. 34-37.

¹² Robert Merrihew Adams (1985), Judith Andre (1983), Margaret Urban Walker (1991), and Bernard Williams (1981, 1993) are notable for taking such a position. Brynmor Browne (1992) is open to the possibility of moral luck, but contends that if it does exist, we ought to at least partially revise our moral practices.

she performs an action that is morally indifferent.”¹³ Pereboom asserts that there are some situations where an agent can be the author of an action where said action is neither good nor bad, and thus the agent would neither be praiseworthy or blameworthy for it. What makes an agent morally responsible is that there is a morally relevant relationship between her and what she is morally responsible for. For the libertarian incompatibilist, as for Aristotle, this relationship is an ownership relationship – a moral agent owns her actions because she, and she alone, brings them about.

When I say that an agent is *morally responsible* for something, I mean that agent’s *moral record*, or moral history, is affected by that thing. One’s moral record is a list of that agent’s moral responsibility, praiseworthiness, and blameworthiness. To be *praiseworthy* is to have your moral record affected in some objective, intrinsically positive way; to be *blameworthy* is to have your record affected in some intrinsically negative way. By “moral record,” I mean to pick out roughly what Michael Zimmerman calls a “*moral ledger*” (2002, 555, my emphasis). For our purposes, these terms are interchangeable, however I prefer the term “moral record” because I believe Zimmerman’s term, “moral ledger,” is inadvertently misleading. I believe it evokes a particular method for determining what I call one’s *total moral worth*, by which I mean something like an ultimate moral assessment of a person, akin to the sort of moral assessment many religions believe awaits us after death. I believe the term “moral ledger” inadvertently implies that to calculate one’s total moral worth, one might simply add up the entries in one’s moral record and arrive at some figure that represents one’s total moral worth.¹⁴ If such is the case, luck seems to play a role in determining one’s total moral worth. The following case illustrates this:

¹³ See Pereboom, 2005, page 18.

¹⁴ Note that Zimmerman never uses the term “moral ledger” in this way.

Chris and Kris are virtuous moral agents¹⁵ whose lives are identical in every morally relevant way up until time t , at which point, by chance, Kris dies and Chris continues to live a virtuous life.

If one's total moral worth is calculated solely by tabulating the entries in one's moral record, Chris's total moral worth, given her greater number of virtuous acts, would be higher than Kris's, and is so because of the contingent fact that Kris dies before Chris. This account is *prima facie* inconsistent with our commonsense moral intuitions, and I reject such an account. There are various quick fixes to this problem (One could divide moral responsibility over the chances to do good, or given a deterministic world we should calculate the total of Chris's actual and potential lives, and Kris's actual and potential lives¹⁶, etc.), but each has their own problems and there are no *obvious* solutions. I wish to remain agnostic about the concept of total moral worth, as it is outside the scope of this dissertation. In chapter 5, though, I will return to this topic briefly, and argue that whatever concept of total moral worth one may mistake for a problem of moral luck ought to be revised.

Moral luck exists if and only if one is morally responsible for something that is a matter of luck, and outside of their control. In other words, moral luck exists if and only if luck plays a role in the contents of an entry in one's moral record, *not* if luck plays a role in the number of entries in one's moral record. Let us call luck in regards to number of entries in our moral record "*record luck*" to distinguish it from moral luck. Although it is uncontroversially true that the total number of entries in our moral record are a matter of luck, it is not at all clear that the *number* of entries in one's moral record contributes to how blameworthy or praiseworthy one is for what they've done. For more on this, see 2.3.2.2.1 below.

¹⁵ By "virtuous moral agents" I mean morally praiseworthy agents whose actions are entirely praiseworthy.

¹⁶ This is similar to Zimmerman's concept of responsibility *tout court* that I discuss in below.

In the moral luck debate, philosophers tend to use the terms “luck” and “chance” interchangeably. I think this is a mistake, and make the following distinction: Something is a matter of *chance* if and only if no agent has control over that thing, while something is a matter of *luck* if and only if the agent in question has no control over that thing. For example, say a coworker gets you a present for your birthday. What you get it outside of your control – a matter of luck for you; but surely it’s not a matter of chance because it was within your coworker’s control. For your coworker it is neither a matter of luck or chance. Note that this isn’t to say that you don’t exert *some* influence over what your coworker gets you – you can drop hints, make it impossible for her to get certain items, etc. But, ultimately, what she gets you isn’t *up to you*, and thus it is outside of your control in the morally relevant sense.

The final terminological distinction I make here is between kinds of control. Even philosophers who believe the control principle is true have different beliefs about what kind of control it requires for moral responsibility. I believe it is important to distinguish between three senses of the term that have been the focus of philosophical inquiry – partial control, complete control simpliciter, and qualified complete control. You can be said to have *partial control* over something if you played some role in that thing’s coming about, or not coming about, or in how that thing occurs or doesn’t occur. In the case above, you may have partial control over what your coworker gets you insofar as she might be influenced in one way or another by what you say, or what do you, or insofar as you limit her choices (such as by going crazy and buying up all of the flowers in the tri-state area to make sure that she cannot buy you flowers).

At best, human beings only have partial control over the consequences of their actions – for example, we might choose to jump into the lake with the goal of saving the drowning child, but whether you can succeed in saving the child (or even in jumping into the lake) is in a

nontrivial sense outside of your control. You contribute to the event by choosing to act (if you don't actually choose to jump into the lake, then *you don't jump*. Perhaps a muscle spasm causes your body to perform the same actions as if you had chosen to jump, but in a nontrivial sense *you* are not doing the jumping because you didn't choose to do so.), but any number of contingent facts outside of your control may interfere with your success.

In contrast to partial control, there are two senses in which one can be said to have be in complete control. To have *complete control simpliciter*, what Zimmerman calls “unrestricted control”¹⁷ over something is for that something to be up you, and you alone, whether or not that thing occurs, such that there is nothing that it outside of your control that could possibly interfere, *or have interfered*, with that thing. This is the sort of control that an all powerful, all knowing god is said to have over everything (or at least over every non-agent). None of us have complete control *simpliciter* over anything, as to have complete control *simpliciter* over something requires that every fact about whether or not that thing occurs be in your control – from the laws of nature to whether or not you were born to bring it about. We exercise no control whatsoever over the past, including our own birth¹⁸, let alone the laws of nature, and thus cannot have complete control *simpliciter* over anything. Because human beings lack moral responsibility for their own original existence, and one needs to exist to be morally responsible for anything, it is a matter of luck that we are morally responsible for *anything* – that is to say

¹⁷ See Zimmerman 1987.

¹⁸ If time travel is possible, then it may be possible to exert some kind of control over the past, but this control is atypical at best. For example, the main character in Robert A. Heinlein's short story “*–All You Zombies–*”, through various acts of time travel, turns out to be his own mother and father. The story leaves it an open question whether this he/she, through still more time travel and perhaps other science fiction, might turn out to be every person who ever existed in his fictional world, and perhaps even to be the cause of everything in his world, including the world itself. This character would exert a unique kind of control over his own birth, but it is not complete control *simpliciter*, as he/she lacks control over his/her birth *at the time of his/her birth*. This control might be called *retroactive* control, but retroactive control implies the possibility that he/she might be able to change his/her past, leading to temporal paradoxes typical of time travel fiction.

that it's a matter of luck that we have a moral record at all. Luck in regards to our existence, and our continued existence, is record luck, not moral luck.

In contrast to complete control *simpliciter*, one has *qualified complete control* if and only if, at the time in question, given the actual past, what one does is *up to her* and her alone. Complete control *simpliciter* requires that at any given time, you can do anything and that nothing in the past or future could prevent you from having done that thing if you choose – even yourself; in contrast *qualified* complete control requires *only* that you have complete control over what you choose to do from the possibilities available to you at that time. (The possibilities themselves are a matter of luck, but what you choose to do isn't). While complete control *simpliciter* is impossible for beings like us (perhaps even impossible altogether), we take it for granted that we have *qualified* complete control over our own free choices and actions.

It is a matter of record luck whether or not we are given the chance to be morally responsible for anything, but when we are confronted with a situation where something is *up to us*, we have *qualified* complete control. In virtue of this, for the rest of the dissertation I will refer to “*qualified* complete control” as *complete control*. The most intuitive interpretation of the control principle is that moral responsibility requires complete control in this sense. If this kind of control is impossible, then no one can be morally responsible for anything.

2.3.2 Kinds of Moral Luck

Thomas Nagel captures the thrust of the problem of moral luck quite nicely when he asks the “How is it possible to be more or less culpable depending on whether a child gets into the path of one's car, or a bird into the path of one's bullet?” (1976, 143) If the control principle is true, the answer is that it's not possible. The problem is that there seem to be several distinct ways in which luck at least *appears* to play a role in determining our moral responsibility.

Nagel distinguishes between four types of luck that appear to play some role in determining an agent's moral responsibility. The first kind, *resultant luck*, is perhaps the most familiar candidate for moral luck, as it is concerned with how luck plays a role in determining the consequences of our actions. For example, an agent may have complete control over whether or not she attempts to rob a bank, but whether or not she succeeds is ultimately up to factors outside of her control – a matter of resultant luck.

Second, *circumstantial luck* is luck regarding the circumstances one faces, or as Nagel puts it, the “moral tests” that we are confronted with (1976, 145). Nagel says “It may be true of someone that in a dangerous situation he would behave in a cowardly or heroic fashion, but if the situation never arises, he will never have the chance to distinguish or disgrace himself in this way, and his moral record will be different.” Nagel presents the following example where circumstantial luck seems to influence one's moral record: Citizens of Nazi Germany, Nagel contends, had the opportunity to act morally and stand against the Nazi regime, or to act immorally by cooperating in the atrocities the Nazis were involved in. Most are culpable, he claims, for choosing the latter. However, many other people were never subjected to this test, and had they been so subjected, many of them would have behaved just as badly. Because of luck alone these people never had to face this test, and may be differently morally responsible than they would have been had they faced the same test.

Third, *constitutive luck* is luck regarding who one is. Traditionally, constitutive luck has been seen as luck concerned primarily with one's possession of certain character traits. However, recently Dana Nelkin has argued that constitutive luck also covers luck in our genes, our upbringing, and all kinds of environmental influences that contribute to determining who we are (2008). Suppose that one's upbringing causally determines that one have a certain kind of

character, say a vicious character, in such a way that the agent lacks any way to prevent coming to have this kind of character. If agents are blameworthy for having such a vicious character, then their constitutive luck in regards to their character would be a case of moral luck.¹⁹

Fourth, *causal luck*, sometimes called *antecedent luck*, is luck dealing with how our choices are determined or influenced by antecedent circumstances. Causal luck is the most bizarre of the four kinds Nagel identifies. He goes so far as to identify causal luck as the classic problem of free will. However, causal luck doesn't require determinism; even if determinism is false, probabilistic laws may also contribute causally such that our choices are influenced by antecedent circumstances.²⁰ Some philosophers argue that causal luck is redundant and can be completely explained in terms of circumstantial and constitutive luck.²¹

Zimmerman groups the last three kinds of luck together, calling them *situational luck*, as they each have to deal with the situation under which an agent chooses or acts.²² Resultant luck can be summarized as luck that follows one's actions, while situational luck captures luck up to, and including, one's actions. In the following sections I will discuss each kind of luck in turn, and argue that kind of luck is not a legitimate case of moral luck.

2.3.2.1 Resultant Luck

Consider the following example of the problem of resultant luck that Zimmerman adapts from Nagel:

Suppose that George shot at Henry and killed him. Suppose that Georg shot at Henrik in circumstances which were, to the extent possible, exactly like those of George (by which

I mean to include what went on "inside" the protagonists' heads as well as what happened

¹⁹ Robert Merrihew Adams (1985) expressed just such a view.

²⁰ See Pereboom, 2002, 41-54 and Watson, 1982, 9.

²¹ See Latus 2001.

²² The distinction between resultant situational luck plays a central role in Zimmerman's account, as he offers two distinct methods for showing why these kinds of luck doesn't influence one's moral record.

in the “outside” world), except for the fact that Georg's bullet was intercepted by a passing bird (a rather large and solid bird) and Henrik escaped injury. Inasmuch as the bird's flight was not in Georg's control, the thesis that luck is irrelevant to moral responsibility implies that George and Georg are equally morally responsible. This, I believe, is absolutely correct. (2002, 560)

The fact that George is a murderer, while Georg is only an attempted murderer turns on resultant luck alone. Zimmerman stipulates that George and Georg are the same situation – so situational luck is not an issue – and that they act in the same way. Only the results of their actions differ. Because moral responsibility tracks control, and thus is immune to luck, Zimmerman argues that both George and Georg must bear equal moral responsibility for what they've done – their moral ledgers, so to speak, have been equally stained. The odd thing is that we're inclined to say that they've *done* two radically different things. George has killed Henry, while Georg hasn't killed anyone; all he's done is shot a bird! What George has done is *intuitively* far worse than what Georg has done. Surely both George and Georg are morally responsible for pulling the trigger with comparable intent, but George has done something Georg hasn't done – succeeded.

We're inclined to say these two are morally responsible for the consequences of their actions; the puzzle is that if this is true, then they should be *differently* morally responsible. Afterall, Henry's death is *far, far* worse than a bird's being shot. Further complicating the issue is that the results of Georg's actions were accidental, a matter of luck undermining his intention, and thus not only is he morally responsible for something less bad, he is less responsible for it as well!

If George and Georg are differently morally responsible, they are so only because of luck, and thus the control principle is false. However, it is at least *prima facie* intuitively plausible that George and Georg are equally morally responsible despite the fact that their actions have different consequences. If this is true, then George and Georg must be morally responsible for *something other* than the actual consequences of their actions.

It is *prima facie* intuitively plausible to say that George is blameworthy for Henry's death; that his connection to Henry's actual death is what makes him morally responsible. However, the same cannot be said for Georg, as Henrik doesn't actually die. What George and Georg do have in common is they both freely chose to kill their target. All things being equal, this choice is *prima facie* intuitively true that both George and Georg are blameworthy for this choice. If this choice is what they are, properly speaking, morally responsible for, then it makes sense to say that they are equally morally responsible because they have done the same *kind of thing*.

The problem with this account is that the inherent badness of Henry's actual death plays no role in determining George's blameworthiness. Because of this, it is at least *prima facie* inconsistent with our moral terminology – we say that George is blameworthy for Henry's death when, all things being equal, according to this theory Henry could have survived and George would have been equally blameworthy. This is to say that on this account Henry's actual death makes *no difference* to George's moral record, and thus, strictly speaking, it is a mistake to say that George is morally responsible for Henry's death.

One solution²³ to this problem, which I reject, is to say that one is morally responsible for *not only* the actual consequences of one's actions, but the possible consequences of one's actions

²³ This possible solution to the problem of moral luck is one of my own design, but is modeled after Zimmerman's solution to situational luck.

as well. Because Georg *could* have killed Henrik, and George could have missed killing Henry, they are equally morally responsible on this account because although the actual consequences differ, the collective pool of active and merely possible consequences of their actions (by stipulation) match. The problem with this account is that it, too, is vulnerable to luck. George and Georg can be the same in all relevant respects up until, and including, their choice to kill their target. However, unbeknownst to Georg, it may be completely impossible for him to kill his target (say unbeknownst to Georg, Henrik is immortal and bulletproof). If this is the case, on this theory, Georg is less morally responsible than George is because his pool of possible consequences differs from George – and does so solely as a matter of luck.

Both George and Georg lack complete control over whether their respective target dies. Any number of factors – from their grasp of the laws of nature, to the accuracy of their sense perceptions, to third party intervention (such as the bird) may prevent their intentional action from having the desired (or feared) effect. They, like the rest of us, lack complete control over the results of their actions, and because of this, they cannot be morally responsible for the results of their actions according to the traditional interpretation of the control principle.

Michael Zimmerman takes the position that only partial control is required for moral responsibility, and contends that George and Georg are equally morally responsible because they have the same amount of (partial) control over their respective target's deaths. For Zimmerman, moral responsibility is determined by (1) what one has (at least partial) control over, and (2) the degree of control one has over that thing. Zimmerman stipulates that George and Georg have morally equivalent goals (their respective targets' deaths) *and* equivalent control over these goals, and thus – in light of this – it makes sense to say that they are equally morally responsible.

It seems to me there are two distinct problems with Zimmerman's account. First, I am not sure what to make of Zimmerman's contention that Georg has partial control over Henrik's death. Henrik *doesn't die*, and the fact that he doesn't die isn't up to Georg *at all*. Georg does his best to kill Henrik and *fails*. When we say "George is blameworthy for Henry's death", it *seems* as if we are saying that there is something about Henry's *actual death* that tracks back to George and stains his moral record. But there can be no equivalent staining of Georg's moral record, because Henrik doesn't actually die. Second, it seems odd to say that the less control one has over a situation, the less morally responsible one is. Consider the following case:

Georgia hates Henrietta for the same kind of reasons that George hates Henry and Georg hates Henrik. Georgia, however, is always up for a game of chance and attaches her rifle to a slot machine, such that when Georgia pulls the slot machine's lever, the slot machine fires the rifle if and only if it comes up three of a kind. Georgia sits atop a building and when Henrietta is lined up in her sites, Georgia pulls the trigger, the slot machine displays triple cherries, the weapon fires, and Henrietta is shot and killed.

In this case Georgia is in *less control* of her victim's death than George is in control of Henry's. However, intuitively, she is *as morally responsible* as George. In Georgia's case, she is *purposely* less in control of Henrietta's death than George is in control of Henry's death, but this is no excuse. Some might draw a distinction, and claim that Georgia is less morally responsible because she didn't choose to kill Henrietta, but that she merely chose to risk Henrietta's life – but this is a mistake. Georgia chooses to try to kill Henrietta and merely chooses a less reliable way to do so than George chooses to kill Henry. We can, of course,

question Georgia's rationale for choosing a less reliable way to commit murder, but for our purposes her moral responsibility for her choice of weapon is a separate issue.²⁴

Georgia is in less control over whether her victim dies than George is – but, intuitively, they are equally morally responsible for what they have done. George and Georg have a different amount of control than Georgia, but are the same in every *other* relevant way, and yet all three are intuitively equally morally responsible. This suggests that the difference in the partial control they have over their actions does not play a role in determining their actual moral responsibility.

Note, though, that there *is* something that all three have the same amount of control over – all three have complete control over their free choice to kill their respective targets. And just as these agents have the same control over their intentional actions, Zimmerman stipulates that the mental content, or “what went on “inside” the protagonists’ heads”, is equivalent. (2002, 560) In Zimmerman’s cases, he leaves this content unspecified – but surely this content matters. Zimmerman places the reader in the same position as the third party observer – we are ignorant of this content for George *et al.*, but we can guess. We assume that George and the like *wanted* to kill their respective targets; that it isn’t some accident or coincidence, and we do so because there tends to be a connection between our intentions and the outcomes of our actions.

Embodied in the concept of intentional action is the idea that there is a regular connection between one’s intentions and what one intends to bring about.²⁵ Anyone short of the radical skeptic presupposes that there is a regular connection between our beliefs about the world and

²⁴ For example, say Georgia chose to drive to work, and then she chooses to drive her SUV instead of her hybrid. The proper moral evaluation of Georgia’s driving to work is, I think, a separate question from her moral responsibility to drive to work in her SUV instead of her hybrid. Georgia can be completely morally justified in choosing *to* drive to work, while being completely morally blameworthy for her choice of *what* to drive to work.

²⁵ If one didn’t believe there was such a regular connection, contingent or otherwise, they would have no awareness of what their free actions would bring about, and have no basis to choose whether or not to act – thus failing to satisfy Aristotle’s awareness criteria.

how it actually is. It follows from such a belief that there is a similar regular connection between what we believe our actions will bring about and what they will actually bring about. And much like the former regular connection, this latter regular connection can be interrupted by both facts about the world we are ignorant of, and by our own biases and desires.

Given this, there is a predisposition for the ignorant third party who knows only that George shot Henry to assume that George intended to shoot Henry. Henry's death is itself *prima facie* bad, so in the absence of some extenuating circumstances (such as if Henry was attempting to suicide bomb others), we are supposed to conclude that George's intentions are bad. But what makes George's intentions bad *just are* that he intends to bring about Henry's death (without any upside). Henry's actual death is bad, but on this view what makes George blameworthy is that his action is designed to bring about Henry's death.

But surely moral agents aren't solely morally responsible for intentional actions. Take the drunk driver, who accidentally hits a pedestrian on her way home. All things being equal, she doesn't intend to hit the pedestrian, but intuitively she is still blameworthy for it. Although the drunk driver's actions aren't intentional, there is still a morally relevant *epistemic connection* between her choice to drive drunk (or, at least, a would-be driver's choice to drink without taking steps ahead of time to ensure that she won't drive) and her awareness of the possible consequences, including the possibility of a fatal accident. This is to say that when the agent acted, she acted *believing* her action might cause certain outcomes. While she may not have intended these outcomes, she freely and needlessly risks these outcomes through her free actions.

When a moral agent acts intentionally, whether pulling a trigger or taking a drink, there are numerous *epistemic connections* that exist between the agent's choice and her belief-sets and desire-sets. For the remainder of this dissertation, by "epistemic connections" I mean epistemic

connections of this kind. These epistemic connections are *prima facie* morally relevant; we tend to judge the ignorant person who unwittingly caused harm *less* morally blameworthy than the well-informed person who willingly and maliciously caused harm. Indeed, if the ignorant person had no reason to believe her actions would cause harm, we often come to the conclusion that she might not be morally responsible at all. However, when we find out that an apparent accident that causes harm was, in fact, planned out, we tend to judge the agent more morally blameworthy than had it merely been an accident. For example, it is *prima facie* morally true that a drunk driver who hit a pedestrian would be less morally blameworthy than an assassin who uses a car to hit his target.

I contend that agents who act intentionally and bring about a certain result are more morally responsible than agents who act intentionally and merely accidentally bring about the same result because the latter have *fewer* and *weaker* epistemic connections than the former. That is to say that the drunk driver has fewer and weaker epistemic connections to her belief that someone may die as a result of her actions than the assassin does.

Identifying how various epistemic connections contribute to one's moral responsibility is outside the scope of this dissertation. However, it seems to me that there are three *prima facie* epistemic connections that play a substantial role in determining the agent's moral responsibility – 1) what the agent *intends*²⁶ her action to bring about, 2) what the agent *expects* her action to bring about²⁷, and 3) what the action is *designed* to bring about.²⁸ Other epistemic connections

²⁶ By what the agent intends to bring about, I mean to capture only the reasons she chose to act on, not the reasons she had to act on, but didn't chose to act on.

²⁷ Suppose Charlie is a level-headed person who, despite the odds, spends his life savings on lottery tickets, hoping to hit it big. Although Charlie intends to win the lottery, and his actions are designed to do so, it would seem odd to say that he expects to win the lottery. Given that he doesn't expect to win the lottery, his choice is at least *prima facie* morally blameworthy, as he risks his well-being without good reason.

²⁸ Suppose Steve decides to marry Sheila. Before his choice, Steve's friends point out to him that there are two reasons for him to marry Sheila: (1) her family's money will provide financial security, and (2) he will get to spend his life with the person he truly loves. It makes sense to say that Steve can decide to marry Sheila *solely* because he

that affect one's moral responsibility include the connection between the agent's free choice and their beliefs about probabilities, their fears, their attention or inattention, their vigilance or negligence, etc. The moral character of any given epistemic connection is likely self-evident even if the aggregate moral character of any given set of epistemic connections in a single action is not.

In the case of the drunk driver, there is a less of an epistemic connection between her free choice and her belief that she might kill someone than there is between George *et al.*'s free choices and their beliefs that they might kill someone. George *et al.* choose their actions because they are designed to bring about a death that they both intend and expect. In contrast, the drunk driver's action is (poorly) designed, intended, and expected to bring her home safely; and her choice to drink (without taking precautions) is designed, intended, and expected to, say, make her feel better. In each of these actions, the would-be drunk driver exhibits no desire to hit any pedestrians, despite knowingly risking the outcome. There are *far*²⁹ fewer and weaker epistemic connections to the would-be drunk driver's beliefs and desires about killing pedestrians than there epistemic connections between out assassins and their beliefs and desires about killing their respective targets.

wants to spend his life with the person he truly loves, even if Steve *also* desires financial security. If Steve marries Sheila for love, it still makes sense to say that there is an epistemic connection between his marrying Sheila and his desire for financial security, but this epistemic connection played no role in Steve's choice. However, this is not to say that it is a *weak* epistemic connection, Steve believes that his action will result in financial security. It is, however, *weaker* than had Steve chosen to marry Sheila for financial security, or out of a combination of both reasons.

²⁹ It is worth noting that the comparative weakness of the drunk driver's epistemic connections to George *et al.* is somewhat misleading. We tend to believe the drunk driver is *substantially* morally blameworthy despite the weak epistemic connections to her beliefs about the possibility of having an accident on the way home. The severity of her blameworthiness is, in large part, determined not by the weak epistemic connection between her drinking and her expectation of risking a person's life, but rather by the substantially strong epistemic connections between her action and her trivial *akratic* desires to feel better. This is to say that the possible gains of her actions *do not justify* the risk she takes. Her reckless regard for human life, while not the same, is at least *comparable* to the assassin's disregard for life.

The epistemic connections between our free choices and our beliefs and desires help to explain our actions – George pulled the trigger to kill Henry. The third party observer can, from witnessing the result, infer certain intentions, desires, and beliefs from the consequences of our actions. Of course the observer has no direct contact with these epistemic connections behind the action, nor the beliefs and desires of the agent, and can only base their moral judgment on the facts at hand – the results of one’s actions. Of course, our actions are designed to bring about certain results in the world – George acted to bring about Henry’s actual death, a natural evil. The problem is, ultimately, Henry’s death isn’t up to George – it’s a matter of luck whether George succeeds, or fails like Georg. However, *why* he acted, given what he knew, *is* assumed to be within George’s complete control. And this, I contend, is what George is morally responsible for.

What should we make of our resultist language, then? We *say* “George is morally responsible for killing Henry”, despite the fact that we believe that Georg is equally morally responsible despite not killing Henrik. Their moral records are the same, but our commonsense moral language suggests otherwise.

Some philosophers appeal to the concept of derivative moral responsibility to help make sense of our seemingly resultist linguistic practices. To be *derivatively morally responsible* for something is to be morally responsible for some *other* thing and for that other thing to be connected to that something in some nontrivial, morally relevant way. It is uncontroversially true that George is morally responsible for his free choice to shoot Henry (designed and acted on with the goal of bringing about Henry’s death), and given that we assume a regular connection between how we view the world and how it actually is there is a nontrivial connection between George’s action which he envisions will bring about Henry’s death and Henry’s actual death.

Because of this, it makes sense to say that George is derivatively morally responsible for Henry's actual death.

Some philosophers contend that derivative moral responsibility is a *kind* of moral responsibility. I think this is a mistake. To be morally responsible for something is to stand in a relationship to that thing such that if that thing is bad, one is morally blameworthy, and if that thing is good, one is morally praiseworthy. To say that George is morally responsible for Henry's death is generally understood as asserting that he is blameworthy for the natural evil of Henry's death. However, suppose George shoots Henry *not* to end his life, but to prevent Henry engaging in a suicide bombing. If this were the case, then George acts to save other people, and shooting Henry is a means to that end. All things being equal, it makes sense to say that if George shot Henry to save others, then George's action is *prima facie* morally praiseworthy. And it is true that George is morally responsible for Henry's death in this case – at least he is so, insofar as his action was designed to bring about Henry's death (before he could bomb innocent people). But, despite all this, Henry's death is still a natural evil, but it *does not* make sense to attribute that natural evil to (heroic) George. It is *a bad thing* that Henry, *qua* person, died, but George shot Henry *to stop the bombing*, not *to bring about Henry's death*, and George's actions are wholly praiseworthy. George is causally responsible, but not *morally blameworthy*, for the natural evil of Henry's death.

In light of this, derivative moral responsibility is not actual moral responsibility; instead it is a byproduct of moral responsibility. This is the case because the actual consequences of one's actions play no role in determining whether one is praiseworthy or blameworthy; rather they provide the third party observer with evidence of the agent's intentions and beliefs with which to judge one's moral responsibility. Although we are derivatively morally responsible for many of

the consequences of our actions – consequences that are, ultimately a matter of luck – our moral record is not affected by them.

2.3.2.2 Situational Luck

Above I've argued that the actual consequences of our actions are outside of our complete control, and thus we cannot be morally responsible for them. Furthermore, I've argued that one can be causally responsible for, and intend to cause, an objectively bad result (the death of a suicide bomber) and despite causing a natural evil, lack any blameworthiness for doing so.

Now I turn to the problem of situational luck. The question at hand is whether or not luck in the situations we face contributes to our blameworthiness or praiseworthiness. If it does, then the control principle is false. In the following three sections, I will discuss the three kinds of situational luck that are said to be candidates for moral luck – circumstantial luck, constitutive luck, and causal luck. I will argue that although luck does, in fact, play a role in determining what situations we face, it does not play a role in determining what we do in these situations, and *that* is what we are, properly speaking, morally responsible for.

2.3.2.2.1 Circumstantial Luck

Thomas Nagel (1976, 145-147) defines circumstantial luck as luck in the circumstances we face, or more poetically “the things we are called upon to do, the moral tests we face”, which he argues are outside of our control. For example, in a given situation, Nagel explains that an agent might act cowardly or heroic, but that if she hadn't been confronted with that particular circumstance, her moral record would have been different. There are two ways in which the entries in one's moral record can be different – in terms of (1) quantity, and (2) quality.

When moral records differ in terms of quantity of entries, it is a matter of record luck, not moral luck. Furthermore, it is not at all clear that the number of entries in one's moral record

affects how blameworthy or praiseworthy a person is *solely* because of the number of entries.

To illustrate this, consider the following case:

Carol and Karol are virtuous moral agents whose lives are identical in every morally relevant way up until time t , at which point Carol makes a donation of \$100 to a charity. At t Karol makes one hundred donations of \$1 to the same charity.

It is *prima facie* true that Carol is *no less* morally praiseworthy for her single donation than Karol is for her hundred donations. Indeed, if neither is aware of any good reason to make multiple donations, it seems that if there is any difference between them morally it is that Karol is *less* morally praiseworthy than Carol. This is true because, all things being equal, Karol knowingly wastes her time, and the charity's time, by making multiple donations.³⁰

Furthermore, if the reason Karol chose to make multiple donations is to *appear* as if she was more generous than Carol, she is *prima facie* morally blameworthy for this.

Nagel's contention, though, is that if an agent's record differs in quality solely because of the circumstances she faces, then it would be a case of circumstantial moral luck. He presents a case of circumstantial luck where two agents' moral records differ substantially primarily because of luck in the circumstances they face. Nagel's case is as follows:

"Ordinary citizens of Nazi Germany had an opportunity to behave heroically by opposing the regime. They also had an opportunity to behave badly, and most of them are culpable for having failed this test. But it is a test to which the citizens of other countries were not subjected, with the result that even if they, or some of them, would have behaved as badly as the Germans

³⁰ This is based on the assumption that Karol is being wasteful, and being wasteful is *prima facie* morally wrong. If being wasteful is not *prima facie* morally wrong, then Karol is not *prima facie* any more or less blameworthy or praiseworthy than Carol solely based on the number of entries in her moral record. Furthermore, if we assume that there is at best a negligible amount of waste involved in making multiple donations rather than one (perhaps this is because a computer system handles the donations), then there is no *prima facie* rationale for distinguishing between Carol and Karol's donations, morally speaking.

in like circumstances, they simply didn't and therefore are not similarly culpable.” (1976, 145-146)

The difference between a typical German and, say, a typical American in this case is that the German had an opportunity the American didn't – one to act heroically or terribly. Many Germans acted terribly, and we have good reason to think that many Americans, confronted with the same circumstances, would act equally viciously. Yet we hold the vicious Germans morally culpable for their viciousness, but Nagel points out we do not hold the would-be vicious Americans as responsible. If the German's moral record differs from the Americans in part because of the circumstances he faces, then it would be a case of circumstantial moral luck.

Nagel's case is the premier example of a situation where circumstantial luck appears to influence one's moral record. However, the case is misleading in two ways. First, although at first glance the case *appears* to be a case of differing *quality* of entries, upon closer analysis the case itself is just a case of differing entries – the German had *at least one* entry that is not comparable to what the American had. Nagel tells us that this is a test the rest of us never had to face. The American was, if you will, *absent*, for the “moral test” that faced the German. Although we might point to some *other* “moral test” that the American had to face at the same time, it is a *different test* – a test the German was similarly absent from. This makes the case an instance of record luck, not moral luck.

The supposition is that the majority of us are better off, morally, because we lucked into never having to take this test. This leads us to the second misleading aspect of the case – the case turns on there being some objective fact about what the American, or any of us, would do in these same circumstances. This seems to presuppose *universal causal determinism*, the theory that the actual past, coupled with the laws of nature, completely necessitate a unique future.

However universal causal determinism is incompatible with the control principle, as nothing would be *up to us*, rather anything we do would ultimately be a byproduct of something outside of our control that happened long before we were born. Of course, we can *guess* what would have happened, and *guess* that many of us would have failed the moral test and behaved like the Germans actually behaved, and this guess is probably true. However, this is akin to guessing about any undetermined outcome. Suppose that we lived in a universe where coin flips were *actually* uncaused events, but where we've determined through observation that there are about as many coin flips that result in tails showing as heads. To say that many Americans would have acted like the Germans did during World War II would be akin to saying that about 50% of these undetermined coin flips would come up heads. It has no bearing on whether any given coinflip would turn up heads, or whether *any* coin flips would turn up heads.

Suppose, though, that we don't treat Nagel's case as a case of record luck (that the Germans have an entry in their record that the American's don't), but as a case of circumstantial luck – say Jane and Janine live morally equivalent lives until time *t*, at which point they both hear a knock at the door. Jane opens the door to find German soldiers asking for the location of Jewish persons to kill, while Janine opens the door to find Girl Scouts selling cookies. Jane freely sells out her Jewish neighbors, while Janine freely buys a box of cookies. It is at least *prima facie* true that Jane and Janine face moral challenges that differ radically in *quality*, not quantity. Suppose we say that Jane and Janine both fail their moral tests (Jane fails her moral test because she is an accessory to murder, while Janine fails her moral test because she gives in to temptation and purchases something that (1) she can't afford and (2) is bad for her health), surely the entries in their moral records differ *substantially*, and do so *solely* as a matter of luck.

Remember, though, moral luck would occur if and only if an agent is morally responsible for something outside of her control. But it is uncontroversially true that no one is morally responsible for the circumstances one faces.³¹ Jane is no *more morally blameworthy* for opening the door to find German soldiers than Janine is for opening the door to find girl scouts. Rather, Jane is morally blameworthy for the choice that she makes. So, too, is Janine. Both are morally responsible for something that is their doing, and thus the entries in their moral record are not a matter of moral luck.

Perhaps the problem isn't that Jane is given an opportunity that Janine wasn't given – Jane is given an opportunity to kill her neighbors without punishment while Janine cannot kill the Girl Scouts without fear or punishment. The supposition is that Janine is put in a position where there is nothing that she can do that is as bad (or good) as what Jane can do in her situation. Janine is contingently less morally blameworthy than Jane, but had Jane done the right thing (what Nagel stipulates as acting “heroically”, whatever that entails) and Janine done the right thing then Janine would be less morally praiseworthy. Coming to any serious conclusion about how morally blameworthy Jane or Janine are for their actions is outside the scope of this dissertation, but I take it that the intuitive thrust of Nagel's case is that the German is *much* more blameworthy than the American. In this case, we're to believe that Jane is *much* more blameworthy than Janine. The problem is that I haven't stipulated anything about the inherently private mental states of either Jane or Janine; from their actions we're supposed to infer that Jane willed, or at least willfully allowed, the death of her Jewish neighbors, while what Janine merely willfully spent money she couldn't afford to spend on something that is bad for her.

³¹ Of course one can be derivatively morally responsible for the circumstances one faces, but this is just to say that the agent is morally responsible for putting herself in such a situation.

For the sake of argument, let's assume that Jane is as the emotive appeal of the case would have us believe – that is to say that she freely turns in her neighbors knowing what will likely happen to them *because* she wants them to be tortured and killed despite the fact that she realizes they are inherently valuable moral agents.³² It seems at least *prima facie* possible that Janine could have done something as horrible as this at time *t*; for example she could have kidnapped the Girl Scouts, killed them, and buried them in her basement. Suppose, though, Janine knew that she couldn't kill the Girl Scouts and get away with it in the way that Jane would get away with her role in murdering her neighbors. If this is the case, Janine could still have formed an intention to kidnap, kill, and bury in her basement the next person or persons she believed she could without getting caught.

When I say Janine could have formed an intention here, I meant that she could have formed a predisposition to act under a given set of circumstances (in this case when she believed she could do so without getting caught). Although moral agents may constantly revise their intentions over time, to form an intention *just is* to will that what you intend occur. It is, thus, at least *prima facie* plausible that such predispositions to act can causally determine the agent to act in the specified circumstances without giving the agent a chance to reconsider. In virtue of this, forming the intention to kill *just is* choosing to kill. Thus if Janine formed the intention to kill the next group of people who she could get away with killing, she is *prima facie* morally blameworthy to a substantial degree; a degree that is at least *prima facie* as morally bad as Jane's

³² We can, of course, imagine a sufficiently misinformed German who believes all Jewish people are evil demons sent to secretly undermine their power, ruin their lives, and kill them in their sleep. Let us also assume this person is such that she is never put in a position where it is morally acceptable for her to reconsider the validity of her views. For such a person, turning in her Jewish neighbors would be akin to turning in evil demons – because that's what she believes they are. She has no control over how things actually are, she can only make choices based on what she believes. And by stipulation, what she believes is that there are evil demons out to get her. To not turn in said evil demons would be to disrespect her own life, and is morally disgusting.

degree of blameworthiness for freely turning in her Jewish neighbors to the Nazis with the aim of getting them killed.³³

That said, it is at least *prima facie* morally possible that two agents who are morally equivalent up until time *t* can be confronted with situations where the worst possible thing the one agent can do is *far* worse than what the other agent can do. But this problem just is a problem of record luck, not moral luck. It is a situation where luck determines what circumstances one faces, but not how one reacts to these circumstances, and how one reacts *just is* what determines the entry in her moral record, not the circumstances themselves. In the cases of Jane and Janine, what they do is up to themselves, to the agent herself, and *that* is all that the control principle requires.

2.3.2.2 Constitutive Luck

Constitutive luck is, broadly speaking, luck in who we are. Traditionally the inquiry into constitutive luck has been primarily focused on luck in regards to an agent's character traits, although more recently some inquiry has been focused on the role luck plays in determining who we are more broadly construed, including the sorts of agents we are, our biology, our upbringing, and the like.³⁴

The problem of constitutive luck is that it seems as if our character traits, our biology, our upbringing, etc. play a role in how we make our choices, what reasons we act on, and thus it is at least *prima facie* plausible that they play a role in determining how blameworthy or praiseworthy we are. Robert Merrihew Adams, for example, argues that people with bad character traits are morally blameworthy simply because they have these character traits, regardless of how they

³³ Of course many Germans probably engaged in similar actions with far less monstrous intentions than we are ascribing to Jane here. Indeed, in many situations it may very well have been morally acceptable to turn in neighbors to protect one's own family. However, this is clearly not the kind of vicious actions Nagel's case is meant to capture.

³⁴ See Dana Nelkin 2008.

come about, and this view *seems* at least *prima facie* plausible.³⁵ But it is false. Adams equivocates between the natural evil of certain character traits insofar as they tend to bring about bad consequences and being morally responsible for having these character traits.

Some character traits may be causally determined by free choices an agent makes. To the extent that an agent is aware that her actions may bring about a given character trait, it makes sense to say that agent is derivatively morally responsible for that character trait. However, many character traits may be causally determined by external factors outside of their control. If these traits rob an agent of free choice, they make it such that the agent is actually *less* morally responsible for having them, not more. However, so long as an agent's character traits do not completely causally necessitate her to act, then she can, at least in part, be the author of her actions. It should be possible for an agent with a hindering or bad character trait to consistently act in a morally praiseworthy manner – indeed, this manner may, in fact, be more praiseworthy in virtue of the fact that acting in this manner may be more difficult because of one's bad character.³⁶

It makes sense to say that some character traits, biological traits, or other traits are naturally evil or good insofar as they make it more or less likely that an agent will act in certain ways. However, the likelihood of these character traits to cause harm is comparable to the likelihood of certain diseases to cause harm to others. It would be *absurd* to hold sick moral agents morally responsible for simply being sick. Why, then, is having a certain *prima facie* bad character trait any more blameworthy than having a certain *prima facie* bad illness? In both cases we may have *prima facie* good reasons to distance ourselves from the agent in question, or even confine her for the public good, and in this sense I suppose we may treat her *similarly* to an

³⁵ See Adams 1985.

³⁶ Because character traits are, at least in part, inherently private mental states, bad character traits may make it more difficult to accurately judge one's moral responsibility from third party observation.

agent who is blameworthy, but these are also the same steps we might take when dealing with dangerous animals, dangerous weapons, or the like. And there is no question that these things are *not* blameworthy for either their actions, or the causes of their actions.

The idea that we are non-derivatively morally responsible for our character traits is absurd. The idea that we could be morally responsible in any sense for our biology is similarly absurd. Constitutive luck may very well affect whether one has a moral record (perhaps one is born without the biological framework necessary for certain mental events required for moral responsibility), and it may influence how others are inclined to judge one morally, but neither of these makes one any more or less morally praiseworthy or blameworthy for the actions which are in one's control.

It is important to note that moral agents are not morally responsible for the extent to which our character traits and/or biology determines our actions. The thief whose character or biology, coupled with actual circumstances she facts, *determines* that she will steal is *not* morally responsible for her actions; although we can say that her character is (a) to blame (causally speaking), and that it is (b) a natural evil. However, the thief whose character or biology just *makes it more likely* that she will steal, who then steals, *is* morally blameworthy for her actions; although the extent to which she is blameworthy is not the same as the thief without these character traits. She is not less blameworthy because she has less control. She is less blameworthy because she is only responsible for those things over which she has complete control, and she has fewer things over which she has complete control than the thief who has no bad character traits. She is, however, more blameworthy than the nomically necessitated thief, as the latter lacked complete control over her stealing.

An agent's character traits and biology are just part of the circumstances an agent is confronted with. As such, constitutive luck *just is* a matter of circumstantial luck, and as we've seen circumstantial luck isn't a kind of moral luck. At best, it's a kind of record luck.

2.3.2.3 Causal Luck

Nagel contends that the problem of causal luck *just is* the problem of free will – for him, moral agents are, ultimately, causally determined at every step, and thus there is nothing which is really up to us in any robust sense that would satisfy the control principle. Nagel contends that the problem of causal luck defies solution. I think this is a mistake. If universal causal determinism is true at the actual world, then no one can ever be morally responsible for anything – at least no one can be morally responsible in the sense discussed by Aristotle and compatible with the control principle.

However, it is at least *prima facie* plausible that the actual world could be governed by indeterministic laws of nature, such that at any given time there can be multiple possible futures. There are at least two *prima facie* plausible ways in which an agent's actions can be undetermined. First, an agent's actions can be caused randomly, by chance. Indeterminism of this kind of no more consistent with the control principle than determinism is.³⁷

Second, an agent's actions can be caused by the agent herself; this is to say that moral agents can be first causes, or authors, of their actions – actions which can be, at least in theory, based upon prior circumstances or luck, but not caused by them. Many philosophers have challenged the coherence of such a position, but the notion that moral agents can have such non-

³⁷ Some philosophers contend that universal causal determinism is more compatible with the control principle than universal indeterminism of this kind. While I admit universal causal determinism might make the world more predictable, I fail to see how an agent can be any more in control of the state of the universe at its creation (from which in a completely deterministic universe every future event is completely causally determined) than by something undetermined, such as a truly random coin flip or, perhaps, the location of an electron in an atom's electron cloud.

arbitrary control over their actions is at least *prima facie* intuitively plausible. It is *this* kind of control that libertarians believe is necessary and sufficient for moral responsibility, and *this* kind of control that both Aristotle and the control principle seem to require for moral responsibility.

Chapter 3: The Principle of Alternate Possibilities

3.1 Introduction

There is something genuinely wrongheaded about the idea that someone can be morally responsible for something outside of her control. The idea that moral responsibility requires control is so central to the concept that to even consider responsibility without control is to contemplate *something else*.³⁸ Despite consensus on the validity of the control principle, though, there is much disagreement about what *kind* of control it requires. According to Aristotle's control condition, to be morally responsible for her actions, an agent must be the cause of her actions in the right way. All too often, at this point, the debate shifts from questions about the kind of control that is *prima facie* intuitively required for moral responsibility, to questions about what kind of causation is possible at the actual world.

The rationale for the shift is simple – we are first and foremost interested in facts about the actual world, our interest in other possible worlds is largely only academic. If we take our moral intuitions and beliefs to be indicative of facts about the world, then we are committed to the following two propositions: (1) moral responsibility is possible, and (2) *right now* there exist moral agents who are capable of being morally responsible for things, many of whom are actually morally responsible, and *we are agents of this kind*. If (1) is true, then the causation required for an agent to be morally responsible must be possible, and if (2) is true, it must occur at the actual world.

³⁸ For example, revisionists about moral responsibility suggest that moral responsibility, of the kind discussed by Aristotle, Kant, and the like, is impossible and/or incoherent. In light of this, they contend we ought to revise our definition of the term to be applicable to the actual world. I will deal with questions of the coherence of libertarian theories of moral responsibility in the next chapter.

The problem with this shift of focus is twofold – (i) it is at least *theoretically possible* that we are not actually moral agents, and thus are not morally responsible for anything³⁹, and if this were the case, (2) is false, and our commitment to the truth of (2) needs to be explained in some way other than by the truth of (2); (ii) despite the long and sordid history of philosophical debate surrounding compatibilism and incompatibilism, we are no closer to a theory of causation that is supported by the empirical evidence of our best scientific theories than when Hume defined causation as mere “constant conjunction” of two objects.⁴⁰ In light of this, shifting focus from the ill-defined concept of control to the more mysterious concept of causation seems only to confuse the issue, and invite disaster.

Due to the apparently intractable nature of the debate over the proper interpretation of CP, a separate, but equally *prima facie* intuitively plausible, principle is often treated as a surrogate for CP – the principle of alternate possibilities⁴¹:

The principle of alternate possibilities: For a moral agent to be morally responsible for some action requires that she could have done otherwise. (AP from now on.)

While the control principle articulates necessary and sufficient conditions for moral responsibility, the principle of alternate possibilities only articulates a necessary condition for moral responsibility – to be morally responsible for something, one needs *at least* to have the alternate possibilities, or the ability to do otherwise. When a moral agent is confronted with a

³⁹ For example, many incompatibilists ascribe to the theory that moral agents can be first-causes, or uncaused-causes, of their actions. It is plausible that we have mistaken ourselves for these kinds of agents when really, as the determinist claims, we are completely causally determined to act by antecedent circumstances. It is another question altogether whether agents of the kind we believe ourselves to be actually exist.

⁴⁰ See Hoefer 2008, Hume, 1999, pg. 121.

⁴¹ CP is supposed to articulate the necessary and sufficient conditions for moral responsibility, while the principle of alternate possibilities merely articulates necessary conditions for moral responsibility. This is to say that not everything that has alternate possibilities is a moral agent. For example, some theories suggest that an electron’s path in the electron cloud surrounding an atom is undetermined, such that an electron has multiple possible locations where it can move to given the actual past. If this is the case, it makes sense to say that the electron had alternate possibilities, or could have done otherwise, but it is clearly not morally responsible for where it goes.

situation where she *literally* has no choice about what she does, she is not morally responsible for what she does because what she does wasn't *up to her* in any relevant sense.

The principle of alternate possibilities allows us to distinguish between moral agents acting *qua* moral agents, and moral agents who are acted upon, or caused to act. For example, AP explains our intuitions about the following pair of cases:

The Diving Cases:

At time t John is standing atop a diving board and spots Joan near the pool below. He decides to jump into the pool with the intention of splashing Joan, jumps, and succeeds in splashing Joan.

At time t Jon is standing atop a different diving board and notices Joann near the pool below. Before Jon can choose what to do, Jack sneaks up behind him and pushes him off the board into the pool. When Jon falls into the pool, Joann is splashed as a result.

The difference between John and Jon is that the former had a choice, and freely chose not only to jump into the pool, but to do so to splash Joan, while the latter had no say in what he did. If John is capable of the sort of action-authorship CP requires for moral responsibility, it's up to him whether or not he would jump at t , and for what reasons. But this is to say that he could have done otherwise, he could have chosen not to jump at t , or he could have chosen to jump at t for different reasons. Jon lacks this control, and this is at least signaled by his inability to do otherwise.

Traditionally, the principle of alternate possibilities has shared the same sort of near-universal acceptance of CP... at least until some forty years ago in his now infamous paper "Alternate Possibilities and Moral Responsibility" Harry Frankfurt broke with tradition and

argued that AP is false.⁴² Frankfurt constructs a case in which he believes he has shown that an agent lacks alternate possibilities, but despite this it is at least *prima facie* true that he is morally responsible for what he has done. If Frankfurt's case holds up to scrutiny, not only AP false, but any meaningful account of CP is as well.

In this chapter I will show that Frankfurt fails to show that AP is false. Following this, I look at several recent variations of Frankfurt-style cases and show that they, too, fail to demonstrate the falsity AP.

3.1 Frankfurt Cases

Frankfurt sets out to construct a case where an agent is *prima facie* morally responsible despite lacking alternate possibilities. I will call all cases constructed with this goal "*Frankfurt cases*". Let any Frankfurt case that succeeds in demonstrating that an agent can be morally responsible for what they've done despite lacking alternate possibilities a "*true Frankfurt case*". A true Frankfurt case would be a counterexample to AP, and demonstrate its falsity. All (honestly presented) Frankfurt cases are *meant* to be true Frankfurt cases, but without exception the prominent Frankfurt cases fail to demonstrate the falsity of AP. I believe there are *no* true Frankfurt cases. Below I will discuss Frankfurt's original cases, as well as two recent variations, and I will show that each of these cases fails to be a true counterexample to AP.

3.3.1 The Original Frankfurt Case

⁴² Frankfurt, Harry G., 1969, "Alternate Possibilities and Moral Responsibility", *The Journal of Philosophy*, Vol. 66, No. 23: 829-839.

Frankfurt's infamous case is meant to be a situation where an agent – Jones₄⁴³ – is morally responsible for his action, despite the fact that it is *prima facie* clear that he could not have done otherwise:

Suppose someone – Black, let us say – wants Jones₄ to perform a certain action. Black is prepared to go to considerable lengths to get his way, but he prefers to avoid showing his hand unnecessarily. So he waits until Jones₄ is about to make up his mind what to do, and he does nothing unless it is clear to him (Black is an excellent judge of such things) that Jones₄ is going to decide to do something other than what he wants him to do. If it does become clear that Jones₄ is going to decide to do something else, Black takes effective steps to ensure that Jones₄ decides to do, and that he does do, what he wants him to do. Whatever Jones₄'s initial preferences and inclinations, then, Black will have his way. (835)

Now suppose that Black never has to show his hand because Jones₄, for reasons of his own, decides to perform and does perform the very action Black wants him to perform. In that case, it seems clear, Jones₄ will bear precisely the same moral responsibility for what he does as he would have borne if Black had not been ready to take steps to ensure that he do it. It would be quite unreasonable to excuse Jones₄ for his action, or to withhold the praise to which it would normally entitle him, on the basis of the fact that he could not have done otherwise. (836)

There is a lot to unfurl here, but before we tackle that monster we should be clear about what Frankfurt believes this case shows. Frankfurt says, “[The principle of alternate

⁴³ Frankfurt begins “Alternate Possibilities and Moral Responsibility” by arguing that the coercion principle is false. He offers three cases, featuring Jones₁, Jones₂, and Jones₃. Frankfurt believes his Jones₃ case to be an example where Jones₃ is morally responsible, despite being coerced, and the Jones₄ is a variation of this case.

possibility's] exact meaning is a subject of controversy, particularly concerning whether someone who accepts it is thereby committed to believing that moral responsibility and determinism are incompatible. Practically no one, however, seems inclined to deny or even to question that the principle of alternate possibilities (construed in some way or other) is true.” But Frankfurt believes that all (charitable) interpretations of AP are false, regardless of the (charitable) sense of “could have done otherwise” that AP is concerned with.

However I think there is at least *one* reasonably charitable interpretation of AP that even a cursory interpretation of Frankfurt's case in no way casts doubt upon:

The weak principle of alternate possibilities: A prerequisite for a moral agent's being morally responsible for acting in such and such a way is that, for that agent, there is some imaginable situation where she would or could have done otherwise. (WAP from now on.)

It is utterly uncontroversial that in Frankfurt's case, Jones₄, *sans* Black's involvement, would have alternate possibilities. It is also utterly uncontroversial that Jones₄, lacking any indication of Black's presence, *believes* that he has alternate possibilities (in whatever sense he believes is required for moral responsibility). Furthermore, Black – an excellent judge of such things – believed that it was an open question what Jones₄ would decide to do. As such, this case isn't a counterexample to WAP, and Frankfurt hasn't shown that *any* reasonably charitable interpretation of AP is false. I believe that WAP is *far too inclusive* to accurately capture our intuitions concerning alternate possibilities⁴⁴ – after all, I can imagine, in *some* sense, devoting my life to curing cancer and magically discovering a cure, but the mere fact I can imagine myself doing so doesn't seem relevant *at all* to whether or not I am blameworthy for not curing cancer.

⁴⁴ Note that WAP, unlike AP, is remarkably consistent with compatibilism.

The rest of this chapter will focus on more traditional, commonsense interpretations of AP, and I will argue that Frankfurt's case fails to be a true counterexample to AP.

For the Black case to be a true Frankfurt case, it must be the case that (1) Jones₄ is morally responsible for something, and (2) that he couldn't have done otherwise than to do that thing. Frankfurt believes (1) is self-evident in this case – it is supposed to be *prima facie* true that Jones₄ is morally responsible for what he does because he does so freely, *believing* that he could have done otherwise. Black's involvement is supposed to secure the truth of (2). The biggest problem with the Black case is that it's not at all clear how Black's involvement *makes it the case* that Jones₄ can't do otherwise.⁴⁵

Frankfurt tells us that the exact meaning of the principle of alternate possibilities is a “subject of controversy” (829), and the terms “alternate possibilities” and “could have done otherwise” are no less mysterious. However, to understand how Black makes it such that Jones₄ can't do otherwise, we must first give some tentative account of what alternate possibilities are (such that we know whether or not Black effectively cut them off in the case).

The most commonsense interpretation of what it is for you to have alternate possibilities is for you to have two or more different possible actions (or sets of actions) *actually* open to you, given the action past, when you chose to act. To have had alternate possibilities, then, is to have had the ability to do not only what you actually did, but to have had the ability to do something else than what you actually did, and for you not to have done so.

⁴⁵ Frankfurt contends that his example is flexible enough to allow the removal of Black in favor of natural forces of a machine. David Blumenfeld, 1971, offers an alternative, replacing Black's machinations for a form of radiation that triggers when Jones blushes. If Jones were to consider doing other than the radiation would have him do, he blushes. Blumenfeld's example differs from Frankfurt's in two ways – trivially, it replaces the overly theatrical Black with a bizarre natural event – no problems there. Second, rather than intervene *only* when Jones is about to choose otherwise, the radiation intervenes whenever Jones even considers choosing otherwise. This deviation, I think, only makes it more susceptible to the Kane/Widerker objection I discuss below.

There are other, less commonsense, accounts of alternate possibilities, most prominent of which is the necessitarian compatibilist position that states that one has alternate possibilities if and only if, had the past been different in some relevant way, one would have been causally determined to do otherwise. On *this* account of alternate possibilities, a murderer who is completely causally determined to murder has alternate possibilities if, had the past been different, she would have done otherwise. The problem with this account is twofold – first, this would be a case a situational moral luck – what makes such a murderer morally blameworthy is a matter of luck, how the actual past causally necessitates her to act, and is thus outside of her control. Second, on this account, *everyone* has alternate possibilities all the time. This is to say that there would be no instances where one lacked moral responsibility because they lacked alternate possibilities, because in every single case, had the actual past been different in some way or another, one would have done otherwise.

If the necessitarian account of alternate possibilities is correct, it is never the case that anyone lacks alternate possibilities for anything. But this is, to a large degree, inconsistent with the commonsense interpretation of AP, and thus I will focus exclusively on the more commonsense interpretation of alternate possibilities.

Frankfurt believes that *every* (charitable) interpretation of AP is false, thus for the Black case to do what Frankfurt believes it does – provide a counterexample to AP – it must show that Jones₄ had no alternate possibilities in *any* charitable account of alternate possibilities.⁴⁶ In what follows I will show that there is no coherent interpretation of the Black case in which Jones both (a) lacks alternate possibilities and (b) is morally responsible for what he has done.

⁴⁶ Note that in Frankfurt's case, Jones has alternate possibilities in the necessitarian sense, as had things been differently – say, had Black not been involved, and had Jones been causally necessitated to choose differently – he would have done otherwise. However, the necessitarian sense of alternate possibilities is far from a charitable interpretation of the kind of alternate possibilities AP requires, and thus should not count as evidence against Frankfurt's case as a counterexample to AP.

Frankfurt believes that in the Black case, Black's potential involvement makes it the case that Jones doesn't have alternate possibilities; however given how the case is presented it's not at all clear how Black's potential involvement does this. In the next three sections I will discuss three interpretations of the Black case, and I argue that each of these interpretations fails to show that AP is false.

3.3.1.1 Black Case Version 1: The Tell

The first version of the case I will discuss is a version derived from comments Frankfurt makes in a footnote; he says, "The assumption that Black can predict what Jones₄ will decide to do does not beg the question of determinism" (835). Frankfurt reasons that, perhaps, Black has studied Jones and has learned that Jones tends to exhibit a certain tell – a twitch – before deciding how to act in cases like this, and that in the past the twitch has always signified that he would act in one way rather than the other, and that this twitch or lack thereof would give Black a good indication of what Jones was about to do.

Version 1: The Tell

Black wants Jones to perform some task A. He has the ability to force Jones to choose to A if he hasn't already chosen to $\sim A$, but would rather Jones freely choose to A than be forced to choose A. Black is familiar with Jones and believes that if Jones is going to choose to $\sim A$ he will exhibit a noticeable twitch at some point *before* he decides. As it so happens, Jones exhibits no twitch, and does what Black wants him to do: A. Jones is completely unaware of Black.

In this case, Frankfurt would contend that (1) Jones is uncontroversially morally responsible for his Aing, where to A is some action like stealing Smith's car, killing Smith, or jaywalking, and

(2) Jones lacked alternate possibilities because if he were going to choose to do otherwise, Black would have noticed the twitch and stepped in before his choice to make sure that he chose to A.

Robert Kane and David Widerker have independently raised an objection to this interpretation of the Black case that has come to be known as the Kane/Widerker objection.⁴⁷ The objection is as follows: Either (A) Jones's twitch is causally related to Jones's decision by some deterministic process involved in his making a choice, such that either (i) Jones chooses *before* he would make the twitch, in such a way that if Jones twitches, he has already chosen to $\sim A$, and thus Black is too late to intervene, or (ii) both the twitch and Jones's choice are wholly causally determined by some prior determining factor; or (B) Jones's twitch is not causally related to Jones's decision, and thus any perceived connection on the part of Black is just coincidence.

If (A)(i), then if Black were to wait for a twitch to intervene, he would be too late to intervene before Jones made his choice and thus (2) is false – Jones could have chosen to $\sim A$, and as it is too late for Black to interfere, he would have succeeded in $\sim A$ ing. If (A)(ii), Kane and Widerker contend that incompatibilists – those who believe determinism is incompatible with moral responsibility – would *not* have the intuition that Jones is morally responsible for his action, and thus at least incompatibilists have good reason to believe that (1) is false in this case. The elegance of the Black case is that Jones is supposed to be uncontroversially morally responsible for his action – but in (A)(ii) this isn't the case. If (B), then Jones's twitch or lack of a twitch is not causally related to his choice, despite Black's belief to the contrary, and thus Black cannot ensure that Jones As simply by waiting for a twitch because Jones can $\sim A$ despite his not twitching, and thus (2) is false.

⁴⁷ See Kane 1985, 1996b; Widerker 1995.

The Kane/Widerker objection shows that there is no way to interpret this version of the Black case where we're inclined to believe that both (1) and (2) are true – that is to say that there is no way to interpret it such that Jones is uncontroversially morally responsible *and* Black has the ability to make sure that Jones doesn't do otherwise.

3.3.1.2 Black Case Version 2: Reprogramming

The problem with the above case is that if Frankfurt doesn't beg the question about determinism, and the twitch is not causally connected to the choice, then Black has no way to ensure Jones will A if he can only act before Jones's free choice. However, Frankfurt seems committed to the proposition that Jones can't do other than actually Aing, *not* that he can't do other than choosing to A. In the following version of the case, if Black intervenes, he does so only after Jones's choice to $\sim A$, effectively sidestepping questions of how Black would predict Jones's choice.

Version 2: Reprogramming

Black wants Jones to perform some task A. He has the ability to make Jones A, but would rather Jones A because he freely chooses to. Black has a sixth sense that allows him to know what other people have chosen to do. Black knows that between Jones's choosing to $\sim A$ and his $\sim A$ ing, there would be time for him to reprogram Jones to believe that he had chosen to A such that Jones will A instead of $\sim A$. As it so happens, Jones chooses to A, and Black does nothing. Jones is completely unaware of Black.

If this is the right interpretation of the Black case, Frankfurt would contend that (1) Jones is uncontroversially morally responsible for his Aing, and (2) Jones lacked alternate possibilities.

This case allows for two series of events – either (A) Jones freely chooses to A and his free choice causes him to A, or (B) Jones freely chooses to $\sim A$, Black intervenes to make Jones

believe that he had actually chosen to A, and this causes Jones to A. If (A), it is uncontroversially true that Jones is morally responsible for Aing. However, if (B) it seems *equally* uncontroversially true that Jones isn't morally responsible for Aing.

However, Frankfurt's position seems to be that if (A), Jones is morally responsible for the outcome of his free choice – his actual Aing, and that he could not do otherwise because if (B), he would have still Aed, even though he is not morally responsible for having Aed. But this is because if Black intervenes, then, properly speaking, Jones has already done otherwise – he has chosen to $\sim A$, and thus (2) is false – Jones can either choose to A or choose to $\sim A$, and thus he has alternate possibilities.

Frankfurt's position seems to turn on the fact that even though Jones had alternate possibilities to *choosing to A*, Black makes it the case that he doesn't have alternate possibilities to actually Aing; no matter what Jones does, he will A, and if he As because he freely choose to do so, he is morally responsible for Aing. The problem with this account is that Frankfurt seems to think there is no difference between Jones's freely Aing and Jones's being causally determined to A; for him they are the same possibility. But this is bizarre – it's true that (A) and (B) have the same outcome, Jones As, but surely Jones's freely Aing is a different thing than Jones's being forced to A.

If this version of the Black case is a counterexample to anything, it's a counterexample to the following interpretation of AP:

The principle of alternate outcomes: A moral agent is only morally responsible for acting in such and such a way if she could have brought about a different outcome. (AO from now on.)

AO derives whatever *prima facie* intuitive plausibility it has from the fact that in most cases where an agent is morally responsible for something, they could have brought about a different outcome. However, whether we can, in fact, bring about a different outcome is uncontroversially outside of our control. Insofar as AP derives its notoriety for being a surrogate for CP, any interpretation of AP that is consistent with holding someone morally responsible for something over which they lack control is an uncharitable one.

Frankfurt is well aware of the controversy over the exact meaning of AP, so it would be uncharitable to interpret him as constructing the Black case to be a counterexample to this rather uncharitable interpretation of AP. More importantly, even if Frankfurt believed that the best interpretation of AP was AO, he wouldn't need to resort to a bizarre case like the Black case for a counterexample. Consider the following case:

The Smiths Case:

Mr. Smith has fallen over the side of the boat, but Mrs. Smith caught his hand as he fell. Mrs. Smith holds onto her husband, but her grip is beginning to slip. She calls for help, but unbeknownst to her, help will not arrive before the limits of her physical endurance are reached and she is forced to let go. For a while, Mrs. Smith has two options – hold on, or let go. Again, unbeknownst to her, the outcome will be the same either way – at some point t_2 , she would reach her physical limits and be causally determined to let go and Mr. Smith would fall to his doom. However, Mrs. Smith lets go at t_1 , where t_1 is some time before t_2 , and Mr. Smith falls to his doom.

I think it's blatantly obvious that Mrs. Smith is morally responsible for letting her husband fall despite the fact that, unbeknownst to her, the outcome would have been the same had she held on as long as she could. We don't need to posit any Black-like evil demons that

may or may not manipulate Mrs. Smith – and in that respect, the Smiths case would be far more elegant than the Black case to demonstrate the falsity of AO.

One might object that in the Smiths case, there are, in fact, alternate possible outcomes – the limits of Mrs. Smith's physical endurances may mean that she can hold on to some time t_2 , but in the case she actually lets go at some time long before t_2 , t_1 . Either – Mr. Smith falls at t_1 , or Mr. Smith falls at t_2 . If these are two different outcomes, this is not a counterexample to AO. If you have such concerns, consider the following case:

A Second Smiths Case:

Mr. Smith has fallen over the side of the boat, but Mrs. Smith caught his hand as he fell. Mrs. Smith holds onto her husband, but her grip is beginning to slip. She calls for help, but unbeknownst to her, at time t_2 her body will experience an involuntary spasm causally determined by an undetermined event at t_1 that will cause her to let go of her husband *unless*, at t_1 she freely decides to let go of her husband which causes her to freely let go at t_2 , *and* prevents the involuntary spasm at t_2 . Mrs. Smith freely chooses, at t_1 , to let go of her husband.

Regardless of what Mrs. Smith does, there is only one outcome or result– at t_2 , she will let go of her husband's hand. Yet it seems *prima facie* true that Mrs. Smith is morally responsible for letting her husband fall in this case, despite the fact that there were no alternate outcomes. Thus AO is false. This case – or something close to it – is a far better counterexample to AO than the Black case, if only because we don't have to posit the existence of someone like Black. This suggests that Frankfurt doesn't intend his Black case to be a counterexample to AO. AO isn't only false – it's *obviously* false. Charity dictates that Frankfurt believes there is something *more* to alternate possibilities than alternate outcomes, and must

believe that some interpretation of the Black case demonstrates that Jones lacks alternate possibilities, not merely alternate outcomes.

3.3.1.3 Black Case Version 3: Choose Again

In the following version of the Black case, Black sets up a situation where rather than forcing Jones to A, if Jones chooses to $\sim A$, Black sets up a scenario where Jones will keep reconsidering his choice until such time that he chooses to A.

Version 3: Choose Again

Black wants Jones to perform some task A. He would rather Jones freely choose to A without his intervention, but if Jones chooses to $\sim A$, Black will know it thanks to a sixth sense. If Jones chooses to $\sim A$, Black will have enough time to set in motion a series of events that will cause Jones to reconsider his choice to $\sim A$. Furthermore, Black has set up a series of events that escalate the appeal of A such that he is sure that at some point Jones will ultimately choose to A. As it so happens, Jones chooses to A, and Black does nothing. Jones is completely unaware of Black.

If this is a proper way to interpret the Black case, Frankfurt would be asking us to believe (1) Jones is uncontroversially morally responsible for his Aing, and (2) Jones lacked alternate possibilities because we would eventually freely choose to A.

The problem with this case is that it is fairly obvious that Jones does, in fact, have alternate possibilities, he can choose to A on the first attempt, or not. Regardless of what Black does, Jones is uncontroversially morally responsible for his choice here, even if he chooses to $\sim A$ and later changes his mind. Jones has alternate possibilities here in the same sense that Mrs. Smith does in the first Smiths case above – he A at his first opportunity, or A later.

But there is another problem with the case itself – how can Black be sure that Jones will ultimately choose to A without assuming determinism? Suppose that the series of events Black sets up are a series of threats against Jones, such that it would be irrational for Jones to choose to $\sim A$. This alone does not make it the case that Jones has to choose to A (although it may make it the case that Jones *should* choose to A), as surely Jones can act irrationally. To assume Jones is causally determined to act rationally just is to beg the question about determinism, and thus fail to demonstrate that Jones is *uncontroversially* morally responsible.

There seems to be no interpretation of the Black case such that Jones is both uncontroversially morally responsible and lacks alternate possibilities. Why, then, are so many convinced Frankfurt has found a genuine counterexample to AP? I believe this is the result of an equivocation. The Black case makes it easy to equivocate between Jones as a free moral agent making his own choices and Jones as Black's puppet being causally determined to act. It is uncontroversially true that if Jones's choices are up to him, then he is morally responsible for them. But it is equally uncontroversially true that if Black somehow usurps Jones's free will, then Jones, *qua* moral agent does nothing. Thus, Jones does have alternate possibilities – he, *qua* moral agent, can A, or he, *qua* physical body, can A, and these are two distinct things. In the former, Jones acts, in the latter, he is acted upon.

Further complicating the issue is that our commonsense moral linguistic conventions *seem* to suggest that moral agents are at least partially morally responsible for the results of their actions. In the previous chapter, I argued that it only makes sense to say that agents are derivatively morally responsible for the consequences of their actions, that this responsibility is a byproduct of one's moral responsibility for one's free choices, and that derivative moral responsibility in no way contributes to one's moral record. Thus, even though it makes sense to

say that Jones is morally responsible for Aing when he freely As, this responsibility is wholly derivative of his free choice to A, and his actual Aing (or failing to A, for example should some shadowy figure, White, use his near-godlike abilities to ensure Black's plans fail by causing Jones to ~A in much the same way, but better, that Black would cause Jones to A) does not contribute to his moral record.

Frankfurt's goal is to show that Jones lacks alternate possibilities, and yet is uncontroversially morally responsible for what he does. However, in each version of the Black case discussed above, he fails to show this. In the first case, either Jones is completely causally determined to act as he does, and thus not uncontroversially morally responsible for what he does, or he has alternate possibilities. In the second, Jones lacks alternate outcomes, not alternate possibilities. In the third case, the best Frankfurt could hope to show is that Jones lacks alternate outcomes, and but he clearly has alternate possibilities even if he lacks alternate outcomes.

3.3.1.4 Mele and Robb's Frankfurt Case

Many contemporary philosophers realize that the Black case fails to live up to Frankfurt's desires, but believe that Frankfurt has shown them the way to constructing actual counterexamples to AP. Two such cases in particular seem promising –first, a case by Alfred Mele and David Robb⁴⁸ offers a cutting-edge account of Black's mysterious power; second a case by Derk Pereboom⁴⁹ purports to be a situation where although an agent has alternate possibilities, he lacks morally robust alternate possibilities and thus lacks alternate possibilities in the moral sense. Mele and Robb's case is as follows:

Mele and Robb's Frankfurt case:

⁴⁸ See Mele & Robb 1998

⁴⁹ Pereboom, 2000, 2001, 2005, forthcoming

Bob inhabits a world where determinism is false, but where some events can still be deterministically caused. At t_1 Black initiates a deterministic process that will cause Bob to steal Ann's car at t_2 unless Bob freely chooses to steal the car at t_2 . (Black doesn't know, or care, that Bob can freely choose to steal the car at t_2 in such a way that his free choice would supercede the deterministic process.) Bob doesn't know of Black or the deterministic process. As it happens, Bob freely chooses on his own to steal the car at t_2 . Intuitively, Bob is morally responsible for choosing to steal the car at t_2 .

Mele and Robb's case has several exciting improvements over Frankfurt's original case – first Black doesn't care whether Bob chooses on his own, rather he is only concerned with ensuring that Bob does what he wants him to. In this sense, Black loses any overly theatrical interest he has in leaving Bob every opportunity to corrupt himself. Second, the action in question is Bob's *choice* to steal the car, not his actual stealing of the car. If there's anything we are truly morally responsible for, and not derivatively morally responsible for, it's our choices. As such, this example sidesteps issues of whether Frankfurt cases gain their legitimacy by equivocating between moral responsibility and derivatively moral responsibility, completely ignoring both the result sense and external sense of alternate possibilities.

There remains the question of how, exactly, Black's process works – and Mele and Robb offer their own account. They ask us to imagine that Bob's decision-making process mirrors that of a machine that works something like this: Machine M is designed to create different colored widgets. The colors of the widgets are determined by the color of the bb that hits a sensor when the machine is not currently in the process of making a widget. Surrounding the machine are several bb guns that shoot different colored bbs towards the machine's sensor. If two or more bbs hit the sensor at the same time, the machine makes a widget based on the color of the right-

most bb to hit the sensor. No other bb that strikes the sensor at the same time factors into the color of the widget.

Mele and Robb claim “Bob is analogous to M in an important respect. He is physically and psychologically so constituted that if an unconscious deterministic process in his brain and an indeterministic decision-making process of his were to “coincide” at the moment of decision, he would indeterministically decide on his own and the deterministic process would have no effect on his decision. This situation is an analogue of a case in which two bbs of the *same color* simultaneously hit M's receptor (while M is not busy making a widget)” (103-104). Black’s deterministic process is akin to striking the sensor on the left-most side, while the result of Bob’s indeterministic decision-making process if he chooses to steal is akin to striking the sensor on the right-most side.

Thus, Mele and Robb contend that (1) Bob is uncontroversially morally responsible for choosing to steal the car, and (2) he had no alternate possibilities because there was no outcome in which he didn’t chose to steal the car, and he *cannot* choose *not* to steal the car because Black’s deterministic process would cause him to choose to steal the car if he wasn’t already going to do so.

The problem with this case is that it is nothing more than a daunting version of the Second Smiths case. As such, this case is merely a counterexample to AO, not AP. There was nothing Bob could do to change the outcome – he could do nothing but choose to steal the car. But it still seems perfectly correct to say that Bob had alternate possibilities – he either freely chose to steal the car for his own reasons, or he was forced to steal the car by the device.

3.3.1.5 Pereboom’s Frankfurt Case

Derk Pereboom also believes that AP is false, but rather than attempt to come up with a case where the agent has no alternate possibilities, he believes it would be sufficient to come up with a case where the agent lacks *robust* alternate possibilities – a case where the agent can do otherwise, but where the agent would be equally morally responsible *regardless* of which possibility she chooses. Pereboom's case is as follows:

The Tax Evasion Case

Tax Evasion (2): Joe is considering claiming a tax deduction for the registration fee that he paid when he bought a house. He knows that claiming this deduction is illegal, but that he probably won't be caught, and that if he were, he could convincingly plead ignorance. Suppose he has a strong but not always overriding desire to advance his self-interest regardless of its cost to others and even if it involves illegal activity. In addition, the only way that in this situation he could fail to choose to evade taxes is for moral reasons, of which he is aware. He could not, for example, fail to choose to evade taxes for no reason or simply on a whim. Moreover, it is causally necessary for his failing to choose to evade taxes in this situation that he attain a certain level of attentiveness to moral reasons. Joe can secure this level of attentiveness voluntarily. However, his attaining this level of attentiveness is not causally sufficient for his failing to choose to evade taxes. If he were to attain this level of attentiveness, he could, exercising his libertarian free will, either choose to evade taxes or refrain from so choosing (without the intervener's device in place). However, to ensure that he will choose to evade taxes, a neuroscientist has, unbeknownst to Joe, implanted a device in his brain, which, were it to sense the requisite level of attentiveness, would electronically stimulate the right neural centers so as to inevitably result in his making this choice. As it happens, Joe does not attain this level of

attentiveness to his moral reasons, and he chooses to evade taxes on his own, while the device remains idle (Pereboom, forthcoming, 9-10).

Pereboom contends that (1) Joe is morally responsible for choosing to cheat on his taxes, and (2) although Joe can do otherwise, he can't do otherwise in a way in which he would be differently morally responsible.

There is, I think, something to be said for Pereboom's focus on robust alternate possibilities. He offers the following account of such robustness: "For an alternative possibility to be relevant *per se* to explaining an agent's moral responsibility for an action it must satisfy the following characterization: she could have willed something other than what she actually willed such that she understood that by willing it she would thereby have been precluded from the moral responsibility she actually has for the action" (Pereboom, 4). Joe is said to lack robust alternate possibilities in this case because no matter how events unfold, he will be morally responsible to the same degree (for cheating).

For Pereboom, Joe lacks robust alternate possibilities because even *sans* the neuroscientist's device if Joe were to reach the requisite level of attentiveness then he still could have chosen to go through with his tax evasion, and had he done so Pereboom believes he would be as morally responsible as he is in the actual case where he never reached this attentiveness. In other words, Pereboom contends that merely reaching the level of attentiveness is not a robust alternate possibility because alone it doesn't mean that Joe's moral responsibility would be any different.

It's not at all clear to me that Joe's reaching an increases level of moral attentiveness would play no role in influencing Joe's moral responsibility, but set this aside for the moment. Rather, focus on claim (1) – that Joe is morally responsible for his choice to cheat. Pereboom

claims that Joe makes his choice at the end of the events described in the case, but if this is the case, then he is entirely causally determined to cheat by factors set up prior to the events of the case, and as such Pereboom begs the question about determinism. Pereboom blatantly states that the only way he can avoid making such a choice is for moral reasons, and although under normal circumstances he may have the power to “overcome” these deterministic forces, the device in his brain makes it the case that he cannot.

Based solely upon the events described in the case, in absence of any information about what happened before the case, we have no basis for holding Joe morally responsible for cheating. He has no say in whether or not he will cheat given the device and the state of affairs at the beginning of the case. In virtue of this, Pereboom can’t lay claim to having universal agreement on (1).

We’re told that Joe chooses to evade his taxes, but this choice is entirely causally determined by circumstances in place before the events described in the case. Joe’s choice, as described, is as much a choice as the thermostat makes when it turns on the heat given certain settings. It is possible that Joe, prior to these events, freely acted in such a way that his being causally determined to cheat is something that he is derivatively morally responsible for (for example, perhaps Joe freely chose to become the sort of person who is predisposed to cheat), but we have no evidence of this. If he did freely choose to act in this way, then is intuitively morally responsible for that choice, but we have no reason to think that Joe didn’t have alternate possibilities for *that* choice. Thus, it seems that either (1) is false, or (2) is.

Regardless Joe’s responsibility, or lack thereof, for his predisposition to cheat, the only control he has during the events described in the tax evasion case are over whether he reaches the level of attentiveness that under normal circumstances would allow him to reconsider whether he

would cheat or not. Properly speaking, there is no choice to cheat that takes place in the example. Rather, the only choice Joe can make during the timeframe described in this case is whether or not he will raise his level of moral awareness. We're supposed to believe that if he did this sort of thing *sans* device, it would allow him the opportunity to make a different choice – a choice about whether to cheat or not. In a normal case, this would be the only way for him to interrupt the causal chain that was set up before the events of the case – necessary, but not sufficient, for him to choose not to cheat, *sans* Black's device, of course.

Note, though, that if Joe believed raising his moral attentiveness would be instrumental in him doing the right thing, either long term or short term, then it is *prima facie* true that Joe *would* be differently morally responsible if he raised his awareness than if he didn't.

Pereboom contends that reaching an increased level of attentiveness wouldn't constitute a robust alternate possibility in virtue of the fact that simply by raising his level of attentiveness, he wouldn't be differently morally responsible because he couldn't be sure that he wouldn't choose to cheat in the decision that (normally) followed from it. In other words, Pereboom's position is that Joe, *sans* device, doesn't have robust alternate possibilities because he *might* still end up cheating, and thus his degree of moral responsibility would be the same.

I think this is a mistake. In the tax evasion case, Joe has control over one thing – raising his attentiveness. But raising his attentiveness in this case is tantamount to *trying* to do the right thing, or, at least, trying to try to do the right thing.

Pereboom's case provides us a unique opportunity to look at the difference between an agent who (a) unreservedly cheats, and (b) tries not to cheat, but ends up cheating anyways. In light of his predisposition to cheat, and the neuroscientist's device, it is not possible for Joe not to cheat. All he can do is take the steps that would normally be necessary for him to choose not

to cheat in a case like this. Is an agent who unreservedly cheats morally equivalent to an agent who tries not to cheat, but ends up cheating anyways? I think not; I think their moral records differ. They may have the same entry for their cheating, but the latter does something the former doesn't – try to do the right thing. As long as Joe believes raising his attentiveness level may lead to him doing the right thing, then he is *prima facie* praiseworthy for doing so.⁵⁰ So, too, is (b) *prima facie* more morally praiseworthy than (a), and Joe has robust alternate possibilities (just not alternate possibilities to cheating). To illustrate this, consider the following case:

The Makeup Test:

Joey failed his math-test but has the chance to make it up this Saturday. Joey believes that he *can* pass the test if he chooses to retake it, and believes that he should, morally speaking, pass the test if he can. Joey has the option to retake the test on Saturday or stay at home. He chooses to retake the test.

It seems to me that just by choosing to retake the test Joey has done something that is *prima facie* morally praiseworthy. We might, of course, wonder about whether Joey actually has the ability to pass the test, and what would happen to his original failing grade if he succeeds on his make-up test, but these are separate questions. Joey's showing up to class is a prerequisite for doing what he believes is right (passing the test he believes he can pass), and if he shows up *for this reason*, then he is praiseworthy in light of it. Even if Joey fails because he, in fact, lacks the knowledge to pass the test, or worse, if his teacher secretly rigged the test to ensure he would fail regardless of how much he actually knew and tried, he is still praiseworthy for showing up.

Similarly, it seems that Joe would be praiseworthy for reaching the requisite level of attentiveness. To use Pereboom's terminology, Joe could have willed something (reaching the

⁵⁰ Joe is only praiseworthy for *trying* to do the right thing here. If he is blameworthy for his predisposition to cheat, he is no less blameworthy in virtue of his doing something *else* praiseworthy. However, it's not at all clear that he is blameworthy for this.

requisite level of attentiveness) such that he understood by willing it he would thereby have been precluded from the moral blameworthiness he actually has for the action (not reaching this level of attentiveness). The moral responsibility in question here is akin to Joey's moral responsibility for "showing up to take the test", *not* the moral responsibility for "failing the first test" or "failing the makeup test."⁵¹

The tax evasion case is especially bizarre. Pereboom either begs the question about determinism, such that it is reasonable to believe (1) is false, or the choice he would hold Joe morally responsible for doesn't occur during the events described in the case, such that we can't say whether (2) is true or not. The only thing Joe is in control of during the events in the example is whether he raises his moral attentiveness level – whatever this is. In light of this, we can ask (1') Is Joe morally responsible for failing to raise his attentiveness level, and (2') did he lack robust alternate possibilities for this. If Joe (falsely) believed that raising his attentiveness level might lead him to do the right thing (without risking further harm), then he is *prima facie* morally responsible for failing to raise his attentiveness level, and thus the answer to (1') is "yes". However, raising his attentiveness level in such a case would be *prima facie* praiseworthy, and thus we have good reason to believe that the answer to (2') is "no" – by raising his attentiveness level, he is differently morally responsible. He is akin to Joey showing up to take the makeup test and failing because the teacher rigged the exam.

3.4 Conclusion

⁵¹ It is, of course, not at all clear how one can be morally responsible for failing a test as long as, while taking the test, one acted responsibly. One can, of course, be morally blameworthy for not studying for the exam ahead of time, for staying up late the night before the test, etc; but so long as one legitimately tries to pass the test, one cannot be blameworthy for failing to pass the test alone. One can, of course, be derivatively morally blameworthy for passing the test insofar as one is blameworthy for not studying or for staying up late, but this is different. That said, were one to goof off during the test, knowing this could lead to them failing the test, one can be blameworthy for this.

According to the control principle, moral agents cannot be morally responsible for something outside of their control, something that is a matter of luck. However, the exact nature of the control required by this principle is rather intractable; in contrast, the nature of alternate possibilities is decidedly less so when conceived as a surrogate for the control principle. The principle of alternate possibilities enjoys a similar intuitive appeal to the control principle despite merely specifying a necessary, but not sufficient condition for moral responsibility.

Frankfurt's attack on the principle has taken on an almost mythic quality, however the Black case shouldn't be taken as evidence that Frankfurt has succeeded in his goal of proving AP false; if anything he has merely demonstrated the falsity of AO, a principle that is inconsistent with CP. In contrast, the principle of alternate possibilities, when conceived of as a surrogate for CP, provides us with robust, necessary conditions for moral responsibility – commonsense conditions that are at least *prima facie* consistent with libertarianism.

In the next chapter I will argue that libertarianism is maximally consistent with both CP and AP. Following that, I argue that the kind of control libertarianism requires for moral responsibility is both coherent and consistent with our best scientific theories.

Chapter 4: What Moral Responsibility Requires

4.1 Introduction

I contend that a version of libertarianism that is consistent with the moral principles I have discussed throughout this dissertation is maximally consistent with our commonsense moral intuitions and beliefs. I have shown that it is maximally consistent with both the control principle and principle of alternate possibilities. In the previous two chapters I have defended these principles against *prima facie* counterexamples, cases of moral luck and Frankfurt-style cases respectively.

In this chapter, however, I will discuss perhaps the most troubling argument against libertarianism – the case against incompatibilism. *Incompatibilism* is the theory that moral responsibility and universal causal determinism are mutually exclusive; that moral agents are not morally responsible for what they have been completely causally determined to do.

Libertarianism is an *incompatibilist*⁵² theory of moral responsibility, and thus according to the libertarian, moral responsibility is incompatible with universal causal determinism.⁵³

In this chapter I offer a defense of libertarian incompatibilist theories from three popular arguments against incompatibilist theories of moral responsibility: (1) what Peter Van Inwagen

⁵² Incompatibilists can be separated into two camps – (1) those who believe no one is actually morally responsible for anything, and (2) – *libertarians* – those who believe moral responsibility is possible and thus universal determinism is false. In (1) there is a further division – (a) *hard determinists*, who believe that (i) moral responsibility is incompatible with determinism, (ii) determinism is true, and thus (iii) contingently no one is ever morally responsible; and (b) *hard incompatibilists*, who believe that (i) moral responsibility is incompatible with *either* determinism or indeterminism, (ii) either determinism or indeterminism is true, and thus (iii) moral responsibility is impossible. The difference between (1)(a) and (1)(b) is that hard determinists believe it is only contingently true that no one is morally responsible, while hard incompatibilists believe that moral responsibility is impossible, and thus it is necessarily true that no one is morally responsible.

⁵³ Most incompatibilists believe that if an agent's actions are completely causally necessitated by events outside of their control, then their actions are not free and they cannot be morally responsible for them. However, some incompatibilists contend that an agent can be morally responsible for causally determined actions as long as at some point in the past, there was indeterminism involved (See Ekstrom 2000, 2003; Mele 1995, 1996, 2006; and Kane 1996b). Theories of this latter kind are inconsistent with the control principle as they open moral responsibility up to luck; rather at best it makes sense to say that agents are *derivatively morally responsible* for such causally determined actions, and are actually morally responsible for their earlier indeterministic actions.

calls the *Mind argument*⁵⁴; (2) the *reason-responsiveness argument*, and (3) David Hume's *endurance argument*. Each of these theories asserts that incompatibilist theories of moral responsibility are inconsistent with a different commonsense moral principle – the control principle, the non-arbitrary principle, and what I call the *endurance principle*. I will show otherwise. I argue that any *prima facie* inconsistency between libertarianism and these principles is a byproduct of one of the ways in which we describe causation, in event-terms. Theories of *event-causation* maintain that the best way in which to understand causation is in terms of events, where events completely causally necessitate other events to occur, where *events* consist of objects, or substances, acting, or being changed, over time. The event-causal theorist maintains that only events cause actions, and thus, although agents may participate in events, they are never the causes of their own actions.

There is a plausible alternative to these event-causal theories; theories of *substance-causation* maintain that substances, not events, cause events, where, again, events consist in the substances being changed over time. Theories of *agent-causation* assert that agent-substances can be a special kind of cause, where agents, such as moral agents, are said to be the authors, or first causes, of their actions. According to agent-causal theories consistent with libertarianism, events, actions, or other substances may influence agent-causes, but do not necessarily determine the agent to act in one way or another. If agent-causation is possible, then agents can be the sort of non-arbitrary, yet undetermined, causes of their own actions necessary for libertarian moral responsibility.

⁵⁴ Van Inwagen explains “The name ‘The *Mind Argument*’ is due to the fact that between 1930 and 1960, versions of the argument appeared regularly in that philosophical journal. One example is R. E. Hobart’s classic essay, “Free Will as Involving Determination and Inconceivable without It” (Hobart 1934). Variants of the *Mind argument* are sometimes called arguments from luck. The thrust of the version discussed here is that anything that arises from an indeterministic process is a matter of chance, and no one has control over what is a matter of chance. Authors who have made such an argument include Almeida and Bernstein 2003, Ekstrom 2000, Haji 1999a, 1999b, 2000a, 2000b, 2000c, and 2001, Mele 1998, 1999a, 1999b, and 2005, and Galen Strawson 1994.

Below I will show that the three common arguments against libertarian theories of moral responsibility *fail* if we understand the world in terms of substance-causation, rather than event-causation. Following this, I offer a defense of agent-causation against five of its most common criticisms. I contend that we have good reason to think that agent-causation, or something like it, offers a plausible theory of causation. Furthermore, to the extent that we believe moral responsibility is possible at the actual world, we have good reason to think that agent-causation, or something very much like it, exists in the actual world.

4.2 Three Objections to Incompatibilist theories of Moral Responsibility

In the following sections I discuss three arguments against libertarian incompatibilism. The first argument, the Mind argument, is the argument that the control required by the control principle is incompatible with an undetermined agent. The second argument, the reason-responsiveness argument, asserts that an undetermined agent's actions are necessarily random or arbitrary, and thus are incompatible with the non-arbitrary principle. Although similar, these two arguments assert radically different things – the first that libertarian agents have no control, the second that their actions are inherently arbitrary, regardless of questions of control. After discussing these two arguments, I turn to Carl Ginet's responses to these arguments. He offers a commonsense interpretation of the problem, but fails to offer a robust solution to the underlying metaphysical problems these objections raise. However, I contend that the metaphysics required by agent-causation makes Ginet's account more plausible.

The third argument I discuss is David Hume's endurance argument against libertarian theories of moral responsibility. Hume contends that it wouldn't make sense to hold agents morally responsible for actions if the cause of their action is fleeting, and that the cause of an undetermined agent's actions are always fleeting – that is to say that nothing of the cause endures

in the agent to make her morally blameworthy or praiseworthy for what she caused. I will show that Hume's compatibilist account of moral responsibility fares little better than what he says about libertarianism, and that the libertarian can offer a better response than he envisions.

4.2.1 The Mind Argument

"*The Mind argument*" is the name Peter Van Inwagen gives to a series of arguments against libertarianism that take the following form: (I) If an agent's choice, decision, or action is not completely causally determined by the laws of nature and antecedent circumstances, then it is determined at random, arbitrarily, or by chance. (II) If it's determined arbitrarily *et al.*, the agent lacks control over what she does. Thus, (III) the agent isn't morally responsible for what she's done. The thrust of this argument is that no choice, decision, or action can be free unless it is completely *nominally necessitated* by antecedent events, where something is nominally necessitated by antecedent events if and only if the antecedent events, coupled with the laws of nature, completely causally determine that something.⁵⁵ Van Inwagen summarizes the argument like this: "If what one does does not follow deterministically from one's previous states, then it is the result of an indeterministic process, and (necessarily) one is unable to determine the outcome of an indeterministic process" (2008, 15-16). The argument is that undetermined agents have *no say* in what they do, no control over their choices, and thus the idea that such agents can be morally responsible is inconsistent with the control principle.

Carl Ginet offers a response to these kinds of arguments, attempting to show that agents whose actions are undetermined are still in control. Ginet characterizes the argument as follows:

⁵⁵ Ginet uses this terminology in his 1989 article "Reasons Explanations of Action: An Incompatibilist Account." The term is instrumentally valuable because it captures the kind of causation typical in a purely deterministic world. Libertarians contend that an agent's actions can be causally determined by the agent without being nominally necessitated

- (1) For any time t and any undetermined event occurring at t : It is not possible for it to have been in anyone's power to determine whether the event or some alternative (undetermined) event instead would occur at t .
- (2) Therefore, for any time t and undetermined action occurring at t : It is not possible for it to have been in the agent's power to determine whether that action or some alternative (undetermined) action instead would occur at t .
- (3) Therefore, it is not possible for a free action to be undetermined. (Ginet, 1989, 19-20, *my italics*)

Ginet admits this argument is valid, but disputes its premises, claiming whatever *prima facie* plausibility these premises have is derived from an ambiguity in sentences like this:

- (A) It was in S's power to determine whether undetermined event E or some alternative undetermined event would occur at t (20).

Ginet claims the ambiguity of (A) depends on ownership of the event in question. When E is not S's action, he contends the natural reading of (A) would be an impossible position – it would require that E be both determined and undetermined, but that there is another reading of (A), one that is coherent, which is as follows:

- (A2) It was in S's power to act in a certain way at t without being nomically necessitated to do so; and it was in S's power to act in some alternative way at t without being nomically necessitated to do so (21, *my italics*).

Here, Ginet contends, because we assume E belongs to S (because it is S's action), it makes perfect sense to say that it was in S's power to perform an undetermined event. He says, "Here each conjunct attributes to S the power to make the case something that in itself is perfectly possible, namely, S's performing an undetermined action. If an undetermined action is possible

then there is no reason to say that an undetermined action cannot be in the agent's power to perform" (21). Ginet continues, "To determine an event is to act in such a way that one's action makes it the case that the event occurs."

Ginet's response to the problem turns on the notion of action-ownership, and in doing so begs the question. Although it is *prima facie* plausible to say that agents own their actions because they are *up to them*, the thrust of the Mind argument is that an agent's undetermined actions *aren't* up to them – they're an arbitrary byproduct of the random indeterministic event that caused them. It is, of course, theoretically possible that the indeterministic event that caused the agent to choose one way over the other is located within the agent herself – for example, some suggest that it might be the result of some undetermined quantum event located inside the brain – but this doesn't make the choice any more up to the agent than had an external quantum event caused her to act in the same way. In virtue of the fact that the choices of undetermined agents are outside of their control, it wouldn't make sense to say that they own their actions. This is to say that according to the Mind argument, undetermined agents lack ownership of their actions because their actions are outside of their control.

Ginet's response rests upon the assumption that moral agents have it within their power to act, and what it is to have the power to act is to control the act itself – to own it. However, to adequately defend incompatibilism from the Mind argument, one must show that an agent's non-nomically necessitated actions are up to them, not the byproduct of some undetermined, inherently arbitrary event outside of their control. In 4.3 below, I argue a theory of agent-causation will allow us to do just this.

4.2.2 The Reason-Responsiveness Argument

Libertarian theories are also often criticized for not being able to account for the role that reasons, desires, and the like play in our decision making processes. A. J. Ayer makes the following argument:

Either it is an accident that I choose to act as I do or it is not. If it is an accident, then it is merely a matter of chance that I did not choose otherwise; and if it is merely a matter of chance that I did not choose otherwise, it is surely irrational to hold me morally responsible for choosing as I did. But if it is not an accident that I choose to do one thing rather than another, then presumably there is some causal explanation of my choice: and in that case we are led back to determinism (1959, 275).

Similarly, J.J.C. Smart contends that “a close approximation to determinism on the macro-level is required for free will” and, thus, moral responsibility (1968, 300-301) – for Smart, to act freely is to act in accordance with reasons. The idea that agents act both from reasons and indeterministically, Smart claims, is unintelligible. John Stuart Mill, too, contends that volitions follow from “moral antecedents”, much as physical effects follow from physical causes. “A volition is a moral effect, which follows the corresponding moral causes as certainly and invariably as physical effects follow their physical causes” (1872, 451-452). He contends that desires, habits, and dispositions, combined with contingent external circumstances, nomically necessitate persons to act.

Ginet argues the opposite – that moral agents can act for reasons without being caused to act by those reasons. To demonstrate this, he asks us to imagine an agent has reasons for doing two mutually exclusive things. In his example, Ginet has to choose how to spend his Saturday afternoons – watching football, or doing philosophical work. He claims that he *wants* to do both but can’t do both, and asks, “Can’t I choose to satisfy one of the motives on some of these

occasions and the other on others of them, without there being a relevant difference in the antecedents on these several occasions?” (1989, 26)

Mill’s answer would be no, he can’t. For Mill in cases of conflicting motives, we do whatever we have the best reasons to do. Ginet is skeptical, and turns to Thomas Reid – Reid claimed that we cannot know whether the will is motivated by the strongest will in such cases unless we know what the strongest will is.⁵⁶ Either the “strongest motive” means the one that won out, and thus is true by definition, or it is determined by some independent criteria. But the only charitable account of Mill’s is the latter – Mill proposes just such a criteria, in typical utilitarian fashion, as meaning “the motive which is strongest in relation to pain and pleasure,” but claims that even if this criterion is not correct, what matters is the repeatability of one’s motives, such that if in a given instance an agent has two conflicting motives, A and B, and the agent was motivated by A, in the *next* instance where the agent was motivated by both A and B, A would again win out (1872, 468-469).

At first glance, this is *prima facie* inconsistent with our intuitions about Ginet’s Saturday afternoons – surely he could choose to watch the game one Saturday, and do what really matters (philosophy) the next. But on closer examination, we have good reason to think that Ginet might weigh his reasons differently between the two days – perhaps, on the second Saturday, his deadline approaching and a boring game on, he chooses to do philosophy. This is to say that it is *prima facie* plausible that his motivations may differ between one Saturday and the next. However, assume – as Ginet does – that his reasons are the same both Saturdays, and yet he chooses football the first, philosophy the second. This alone wouldn’t be enough to show that the strongest motivation doesn’t win out – it is possible that Ginet’s motives for watching football are *exactly the same strength* as his motives for doing philosophy. If so, he may be akin

⁵⁶ Ginet, 1989, discusses a passage from Reid’s 1815 work, quoted from Dworkin ed. (1970) 88-89,

to Buridan's ass – caught between two equally motivating choices. To avoid starving – or simply standing there dumbfounded until one of his motives goes away (say the game ends), moral and non-moral agents alike need *some way* to break the tie. Surely Mill would be open to some arbitrary – perhaps even indeterministic – tiebreakers in such cases. However the existence of indeterministic tie-breakers alone wouldn't be sufficient to satisfy the incompatibilist.

Our intuitions seem to mirror Ginet's contention – that moral agents often have the ability to, and actually do, choose to act from two or more dissimilar sets of reasons, and they *can* do so independent of how they actually rate them, whether morally, practically, or by some other scale. For example, we believe moral agents can, and often do, do the wrong thing knowing it's morally wrong.

Perhaps the most troublesome aspect of Ginet's example is that it bears a striking similarity to an example proposed by Peter Van Inwagen (2000) as an argument *against* incompatibilism. Van Inwagen contends that one, flawed, response to the Mind argument (and one that applies equally to the reason-responsiveness argument) is that terms like “chance” and “random” are best understood in terms of a sequence of events, and thus are not appropriate for individual instances of agent action explanation.

To demonstrate the futility of this argument, he first asks us to imagine that undetermined free acts occur. At $t1$ Alice is put in a difficult position and has to choose between lying and telling the truth. Alice considers both and at $t2$ freely chooses to tell the truth. Next, he asks us to assume that free will is incompatible with determinism, and thus Alice's choice is undetermined. Suppose, he says, that the moment Alice finishes telling the truth, God rewinds the universe back to $t1$, and then lets things go forward again. How, he asks, might things turn

out the second time? Because Alice's choice was undetermined, she can either lie or tell the truth. Her telling the truth the first time around has no bearing on what she would do the second time around. Next, he asks us to consider what would happen if God made Alice choose in this way a thousand times. Although we can't say what would have happened, Van Inwagen thinks we can say what "*probably* would have happened: sometimes Alice would have lied and sometimes she would have told the truth." He asks us to imagine we are watching each successive replay, and asks "Is it not true that as we watch the number of replays increase, we shall become convinced that what will happen in the *next* replay is a matter of chance?" (2000, 14-15).

What is striking about Van Inwagen's case is that it seems to be an idealized variation of Ginet's Saturday afternoon case. In Van Inwagen's case, we're supposed to come to the conclusion that Alice lied sometimes and told the truth other times, however she had exactly the same reasons to do either. Suppose we even stipulated that although Alice had reasons to lie and reasons to tell the truth, that one of these sets of reasons was *substantially* greater than the other. For the libertarian, it must still be possible for Alice to do either, and thus Van Inwagen would contend, that in some run-throughs she would lie, and in some she would tell the truth, and since by stipulation there is no change in her reasons, and her reasons favor one choice rather than the other, the fact that we believe she would in some run-throughs tell the truth, and in others lie, is evidence that her choices are arbitrary. This is to say that although, after the fact, we can point to a set of reasons she had for doing what she did, her actual choice to lie or tell the truth has nothing to do with these sets of reasons – on some run-throughs, we believe she would lie, and on others, tell the truth, and how she chose between acting on one set of reasons or the other is a matter of chance.

Van Inwagen's case makes the implications of Ginet's response all the more damaging – in Van Inwagen's case, we might take the position that (1) Alice always chooses to tell the truth in each run-through, but that (2) nothing about the set of circumstances nomically necessitate Alice to tell the truth. But Ginet's Saturday afternoon case seems to commit him, and perhaps the rest of us, to the opposite conclusion – that we believe we not only *can*, but *do* often act arbitrarily.⁵⁷

It seems as if the libertarian has two possible replies – first, she can bite the bullet and accept that agents can be morally responsible for these actions, even though they are ultimately arbitrary. For Ginet, it's apparently sufficient that an agent's undetermined act simply *appears* as if the agent was the cause.⁵⁸ This would mean accepting that libertarianism is incompatible with the non-arbitrary principle, as well as the majority of our commonsense moral beliefs.

Second, the libertarian can contend that agents can act for reasons, but not be causally determined by them. I believe this reply is maximally consistent with our commonsense intuitions and beliefs, most notably it is consistent with our intuition that moral agents can, and often do, do *irrational* things. Ginet might very well put off doing philosophy to watch the game

⁵⁷ I believe there is an ambiguity in Ginet's case – on the one hand, we're inclined to believe that Ginet can choose to watch the game one Saturday and do philosophy the next despite having, all things being equal, the same reasons on one Saturday as the next. However, unless we postulate that Ginet bangs his head and forgets the previous Saturday, and the entire week, and remains clueless to the passage of time, the fact of the matter is that *things change* between the first and second Saturdays in a way in which they do not change in Van Inwagen's Alice case. Thus, if we're inclined to believe that Alice may consistently choose to tell the truth, but not be nomically necessitated to, we can *also* side with Ginet's case and say that he can do one thing one Saturday, and the other the next because circumstances change. But if this is our interpretation of Ginet's case, it is no longer a counterexample to Mill's theory that we can, in fact, act according to our strongest reasons.

⁵⁸ Randolph Clarke in "Incompatibilist (Nondeterministic) Theories of Free Will," an entry to the Stanford Encyclopedia of Philosophy, says the following of Ginet's view: "On Ginet's view, each basic action is characterized by an actish phenomenal quality, its seeming to the agent as if she is directly making happen the event that is her basic action. Ginet stresses the "as if" nature of this appearance; it does not, he says, literally represent to the agent that she is causing the event in question. And in fact, he holds, it cannot be true that we really are agent-causes of what we do, and it need be the case that any events involving us cause our actions." Although Ginet claims agent-causation is impossible, and that event-causation is the only sort of causation that is possible for moral agents, Ginet's view counts as a noncausal theorist's view in light of the fact that he offers no substantive account of the conditions under which an action has these apparently morally decisive qualities.

despite the fact that, objectively, doing philosophy may be both the morally right thing and the *all-things-considered* practical thing. For this, he would be morally blameworthy in virtue of the fact that he knows the right thing to do (philosophy), knows that it is the right thing, and still chooses to act otherwise for a lesser reason (his desire to watch football).

Similarly it is *prima facie* plausible to say that moral agents have the ability to act arbitrarily – this is to say that Alice might very well make her choice to lie or tell the truth arbitrarily, and if this is how she chooses to make her choice, then we shouldn't be surprised if in some run-throughs, she lies, and in others she tells the truth. However, when she chooses at random, she does *not* choose for *any* reasons.

For the libertarian, to be morally responsible for some action is to own that action – to have that action be up to the agent herself. If agents were nomically necessitated to act according to their strongest reason (regardless of whether their strongest reason is objectively rational), there is no way in which the agent's action belongs to the agent. Intuitively, moral responsibility requires that the agent have the ability to act both rationally, for reasons, *and* irrationally, against reason. Only then can it be *up to the agent* whether she acts rationally or not, whether she freely chooses to do good or wrong, and thus whether or not she is morally praiseworthy or blameworthy.

4.2.3 The Endurance Argument

The third argument I will discuss against libertarian theories of moral responsibility is David Hume's *endurance argument*. In *A Treatise on Human Nature*, Hume argues that the libertarian position is founded upon three misconceptions. First, that agents often experience a false sensation of the ability to do otherwise. Necessity implies the use of force, he contends, and because agents feel no force, they mistakenly believe that they could have done otherwise,

despite the fact that they are causally determined to act by their own desires.⁵⁹ Second, there is a “*false sensation of experience*” of liberty that many libertarians “regard as an argument for its real existence.”⁶⁰ But these assertions merely beg the question. Third, Hume contends that religion unnecessarily biases us towards libertarianism. His attempted refutation of this “misconception” leads us to what I call the endurance argument against libertarianism.

Hume contends the libertarian position is generally better received than the *necessitarian* position – the position that moral responsibility requires determinism – because religion has taken a bizarre interest in the matter.⁶¹ Regardless, he claims, the role that moral praise and blame play in society is clear. Hume claims “’Tis indeed certain, that as all human laws are founded on rewards and punishments, ‘tis suppos’d as a fundamental principle, that these motives have an influence on the mind and both produce the good and prevent the evil actions.”⁶² He contends that the concepts of reward and punishment found in religion, moral discourse, and the law all *presuppose* determinism.

This reasoning is equally solid, when apply’d to *divine* laws, so far as the deity is consider’d as a legislator, and is suppos’d to inflict punishment and bestow rewards with a design to produce obedience. But I also maintain, that even where he acts not in his magisterial capacity, but is regarded as the avenger of crimes merely on account of their odiousness and deformity, not only ‘tis impossible, without necessary connexion of cause and effect in human actions, that punishments cou’d be inflicted compatible with justice and moral equity; but also that it cou’d enter into the thoughts of any reasonable being to inflict them. The constant and universal object of hatred or anger is a person or creature

⁵⁹ David Hume *A Treatise on Human Nature* 2.3.2.1.

⁶⁰ David Hume *A Treatise on Human Nature* 2.3.2.2.

⁶¹ David Hume *A Treatise on Human Nature* 2.3.2.3-4.

⁶² David Hume *A Treatise on Human Nature* 2.3.2.5.

endow'd with thought and consciousness; and when any criminal or injurious actions excite this passion, 'tis only by their relation to the person or connexion with him.⁶³

Here, Hume asserts (1) given that punishment and reward are designed to produce obedience, it would be unreasonable to engage in praise or blame unless we believe they would reliably do so, and (2) that it only makes sense to praise or blame an agent who has some enduring connection to her good or bad actions. Libertarians, I think, have good reason to reject (1) – surely it would be sufficient to praise and blame agents if doing so merely made it more likely they would act as you desire. Furthermore, we often imprison or kill morally blameworthy agents, which tends to reduce their opportunities to repeat offend *regardless* of whether they are libertarian agents. However, (2) has far more bite – surely there must be something of the agent that endures such that it makes sense to hold them morally responsible for their actions – but the libertarian's free choices are fleeting, as their actions can be traced back to a bout of indeterminism, Hume says, not the enduring character traits or desires that necessitarians believe causally determine one's actions. Hume says:

Actions are by their very nature temporary and perishing; and where they proceed not from some cause in the characters and disposition of the person, who perform'd them, they infix not themselves upon him, and can neither redound to his honour, if good, nor infamy, if evil. The action itself maybe blameable; it may be contrary to all the rules of morality and religion: But the person is not responsible for it; and as it proceeded from nothing in him, that is durable or constant, and leaves nothing of that nature behind it, 'tis impossible he can, upon its account, become the object of punishment or vengeance.

⁶³ David Hume *A Treatise on Human Nature* 2.3.2.6.

According to the hypothesis of liberty; therefore, a man is as pure and untainted, after having committed the most horrible crimes, as at the first moment of his birth...⁶⁴

Hume's argument is that holding agents blameworthy or praiseworthy wouldn't make sense if their choices were made indeterministically because there is no enduring character trait to encourage or discourage. In light of the *prima facie* plausibility of Hume's argument, the following principle seems to capture our moral intuitions:

The Endurance Principle: For an agent to be the appropriate object of moral praise and blame for some action A, there must be some enduring feature that persists in the agent that played a causal role in bringing about A.

For Hume, what endures is one's character traits. It makes sense to blame and punish moral agents because in most cases, those agents have character traits that would, given a similar set of circumstances, cause them to act in a similar undesirable manner. So far, so good – the problem with his view is that once an agent successfully abandons her bad character traits, or at the point when it is impossible (or highly unlikely) that external circumstances coupled with those bad character traits will cause the agent to act in an undesirable manner, for Hume it no longer makes sense to hold the agent morally blameworthy. This is to say that, for Hume, reformed agents are *no longer* morally responsible for what they've done in the past because they no longer have the character traits in question.

On Hume's view, the moment that someone's character changes, she is no longer, properly speaking, morally responsible for her actions caused by her now lost character *regardless* of how this change in character arises. Furthermore, on Hume's view, should the world be changed in such a way to make someone's character less problematic, again she is no

⁶⁴ David Hume *A Treatise on Human Nature* 2.3.2.6.

longer morally responsible for what her character, coupled with how the world was, caused her to do because she won't be caused to act in the same way again.

On Hume's theory the murderer who immediately recants is entirely morally blameless, the drug dealer who "just wanted to do it once" is entirely morally blameless, and the thief who succeeds in stealing the one and only item she wanted to steal is *entirely morally blameless*. This is because no amount of punishment will make these agents less likely to repeat offend than they already are. Hume would object to cases like these, contending that the character traits that caused them to act wrongly are still a sufficient threat to hold such agents morally responsible. However, this is an empirical matter, and it is at least *prima facie* plausible that prompts such as getting caught coupled with the threat of punishment *alone* might be sufficient to change one's character, and it is at least *prima facie* plausible that an agent's actions can change the world in such a way that their character is no longer going to cause problems of a similar nature.

Although the endurance principle is *prima facie* consistent with our intuitions about moral responsibility, its implications for the necessitarian are very counterintuitive. Of course, the necessitarian compatibilist would be quick to point out that the libertarian's theory has far greater unintuitive aspects – at least, according to Hume, moral agents can be morally responsible for their actions for the time they retain their (mutable) character traits. Hume believes that it wouldn't make sense to hold a libertarian-moral agent morally responsible because the cause of her actions would be a momentary bout of indeterminism. However, the libertarian maintains that the agent herself is the cause of her actions, and there can be no doubt that such an agent is *at least* as permanent as her character traits.⁶⁵

4.3 Agent-Causation

⁶⁵ The only exception to this would be in transitory agents. For example, suppose that there was some hardwired character trait shared by a human being with multiple personality disorder. If it makes sense to say that each of these personalities is a distinct person, then character traits may outlive persons in this bizarre case.

Hume defines causation in two ways – first, the constant conjunction of two objects, and second, the inference of the mind from one to the other.⁶⁶ Neither offers any robust metaphysical explanation of causation, nor captures our commonsense intuitions about causation. Over time, philosophers have settled on two relatively robust metaphysical theories of causation – The first, *event-causation*, claims that events cause other events, where *events* are complex entities typically said to be comprised of objects, or substances, over time. Moral philosophers who appeal to event-causation to explain an agent's actions have trouble showing how the agent who participates in said events has the control required by the control principle. For event-causal compatibilists, agents play the same *kind* of role as other substances in the event, and thus it makes little sense to say that the agent was in any more control of her actions than the laws of physics, the air surrounding her, etc. In contrast, according to event-causal libertarians, agents might play even less of a role in what they do, as an agent's choices aren't determined by the agent herself, but rather by indeterministic mental events that cause her choice.

According to the second account of causation, *substance-causation*, substances, not events, are causes. *Agent-causation* is a special kind of substance-causation according to which agents are special kinds of substances that can have non-arbitrary control over what they do. For agent-causal theorists, agents are *causa sui*, or self causes. Unlike the agents posited by event-causal theorists, what the agent-causal agents do is, in no uncertain terms, *up to them* insofar as they, and they alone, cause their actions, and their choices are not nomically necessitated by antecedent circumstances or undetermined events.⁶⁷

⁶⁶ David Hume *A Treatise on Human Nature* 2.3.2.4.

⁶⁷ Some Philosophers who have held agent-causal views include George Berkeley (1710), Roderick Chisholm (1966, 1971, 1976a, 1976b, and 1978), Randolph Clarke (1993 and 1996), Alan Donagan (1987), Meghan Griffith (2005 and 2007), Timothy O'Connor (1995, 1996, 2000a, 2000b, 2002, and 2005), Thomas Reid (1788), William Rowe (1991, 2000, 2003, and 2006), Richard Taylor (1966 and 1992), Francisco Suárez, John Thorp (1980), and Michael Zimmerman (1984). Hard incompatibilist Derk Pereboom (2001, 2004, 2005, 2007, forthcoming,

What distinguishes an agent-cause from other substance-causes is that agent-causes are stipulated to be complex, reason-responsive substances that, although not caused by reasons, are capable of acting in accordance, or in defiance, of reason. In virtue of this fact, if agent-causation is possible, agent-causal libertarian theories have a response to the reason-responsiveness argument. In virtue of this special attribute of agent-causes, agent-causal libertarians have the tools they need to answer the challenges of the Mind argument – what an agent does, according to the agent-causal theorist, is uncontroversially *up to the agent*, and not any external fleeting moment of indeterminism, and because the agent-cause is capable of acting for reasons, the agent-cause's actions can be non-arbitrary. (Of course, an agent-cause may also choose to act in an arbitrary manner, but having the capacity to act arbitrarily is different from necessarily acting arbitrarily. Thus an agent who acts arbitrarily can be blamed for acting in this manner because she could have done otherwise.) Thirdly, because agent-causes persist over time, it makes perfect sense to say that they are enduring substances of the kind our intuitions behind the endurance principle requires.

4.3.1 Five Arguments against Agent-causation

An agent-causal libertarian theory of moral responsibility is *prima facie* maximally compatible with the control principle, principle of alternate possibilities, non-arbitrary principle, and endurance principle. However, there has been substantial pushback against the theory. In the following sections I will discuss five grounds on which theories of agent-causation are often challenged. First, some contend that agent-causation is needlessly ontologically complex. Second, I discuss a version of the reason-responsiveness argument similarly aimed at agent-causal libertarians. Third, I discuss a version of the Mind argument levied specifically against

forthcoming 2) contends that agent-causation is required for free will, but that there is sufficient evidence against agent-causation to conclude that we lack free will.

agent-causal libertarians. Fourth, some argue the concept of an indeterministic, but non-arbitrary cause is unintelligible. Fifth, some claim that the science evidence shows that we are contingently not agent-causes.

4.3.1.1 The Ontological Complexity Argument

One objection to agent-causal theories is that they require a needlessly complex ontological commitment – belief in both (non-agent) substance-causation *and* the robust agent-substance-causation necessary for libertarian moral responsibility. This *ontological complexity argument* rests upon *Occam's razor* – the theory that one ought not postulate metaphysical entities beyond those that are necessary to do the job. Roughly, this is the theory that all thing being equal, if two distinct sets of ontological commitments have the same philosophical benefits, the more parsimonious is assumed to be true. A variant of the Occam's razor asserts that increased ontological commitments are only acceptable when accompanied by substantive additional explanatory value. Critics contend either (1) the explanatory power of agent-causal theories can be had with a less burdensome ontological commitment, or (2) the *additional* explanatory power of agent-causal theories is not worth the substantially more burdensome ontological commitment given the alternatives.

While it is possible that another theory may have the same explanatory value as agent-causation, critics who allege (1) tend to point to theories of event-causation – specifically either event-causal deterministic compatibilist theories or event-causal indeterministic incompatibilist theories. In this chapter, we've seen how the latter is incapable of matching up with our intuitions about the control, non-arbitrary, and endurance principles; but event-causal

compatibilist theories fail for the same reasons⁶⁸. In virtue of this, the rest of this section will focus on arguments of type (2).

To determine whether or not agent-causal libertarian theories are worth the additional ontological commitment, there are two questions we must answer: (a) what is the burdensome ontological commitment associated with agent-causation, and (b) is it worth it?

First, some claim agent-causal theories employ what Robert Kane calls “*extra factor*”⁶⁹ strategies; all Kane means by this is that some agent-causal theories are best understood as multiple-causal theories, where substance-causation alone is insufficient to explain causation in all cases, and event-causation is needed to fill in the gaps. Some proponents of multiple-causal theories appeal to mind/body dualism, contending that the mind is not subject to normal physical laws, and here is where agent-substance-causation comes in. Kant contended that free will couldn’t be explained in scientific, physical, or psychological terms, and famously posited a “noumenal self” that is outside of both space and time that can’t be explained in this way. Recently, Sir John Eccles has appealed to what he calls a “transempirical power center” that may account for libertarian free will.⁷⁰ Eccles contends that free will requires indeterminism in the brain. However this indeterminism isn’t sufficient for free will, rather this indeterminacy creates “gaps in nature” that “make room for free will.”⁷¹ These gaps are where he claims a transempirical power center can intervene in the physical world, causing the brain to choose one option over another.

⁶⁸ Event-causal determinist compatibilist theories are inconsistent with the control principle, as agents are no more the causes of their actions than they would be if their choices were determined by momentary bouts of indeterminism. Similarly these theories are inconsistent with the non-arbitrary principle because the agent’s actions are ultimately nomically necessitated by contingent facts about the world. And, finally, these theories are inconsistent with our intuitions that support Hume’s endurance principle, as people are only blameworthy or praiseworthy so long as they are likely to act in a similar manner, which is, ultimately, is a contingent fact about the world independent of the action they are said to be praiseworthy or blameworthy for.

⁶⁹ *My italics*, see Robert Kane 2009

⁷⁰ See Eccles 1970 and Popper and Eccles 1977.

⁷¹ Kane, 1996b pg. 117

According to such theories, most causation is event-causation, while substance-causation – specifically agent-causation – is appealed to only to explain libertarian agency. Such theories are inelegant, given that pure event-causal theories offer nearly the same explanatory value – at the “low cost” of rejecting our moral commitments to principles like the control principle or the principle of alternate possibilities. Is the substantial reduction in ontological commitment worth rejecting these commonsense, deeply held moral principles? I believe that it is *not*.⁷² However, there is an alternative to these kinds of mixed-causation worldviews, one that doesn’t posit the existence of event-causation.

All things being equal, there is nothing inherently more ontologically troubling about substance-causation than event-causation. The key difference is that the latter places the causal force behind temporally fleeting events, while the former places it behind persisting substances that change over time. In virtue of this, consider a world governed wholly by substance-causation, including agent-causation. Make no mistake, a world with agent-substance-causation *is* more ontologically complex than a world with only non-agent-substance-causation; however the ontological distance between a non-agent-substance causal world and a substance-causal world with agent causation is *far less* than the distance between the purely event-causal world and the mixed causal worlds discussed above.

Pure substance-causal theories are often rejected outright because both our commonsense linguistic practices and scientific theories often include talk about events causing other events. If there is no event-causation, critics fear, both our commonsense linguistic practices and scientific theories are wrong, and thus adopting pure substance-causation would undermine more of our

⁷² In chapter 1, I contend that radical interpretation of our beliefs commit us to assuming that our beliefs are (a) largely internally consistent, and (b) accurately represent the world. To reject these moral principles *just is* to assert that our worldview is *largely wrong*, that we are *wrong about most things* and that we are *wrong about those things that concern us the most*. The introduction of such *massive error* into one’s worldview simply to posit one less metaphysical entity is *absurd*.

beliefs than adopting pure event-causation. But this is a mistake. The conception of an event that event-causal theorists appeal to has no correlation in any modern physical scientific theory.⁷³ The practical value of our linguistic practices and scientific theories are not undermined by adopting a pure substance-causal theory – at least no more so than adopting traditional event-causal theories.

Note that substance-causal theorists do not deny that events occur; rather they maintain that events are caused by substances. In reality, the philosophical theories of substance-causation and event-causation are practically identical. For example, for the event-causal theorist, event e_1 causes event e_2 , where e_1 is comprised of substances, s_1 changing to s_1^* over time between t_1 and t_2 . For the substance-causal theorist, e_1 causes s_1 to change to s_1^* between t_1 and t_2 , and s_1^* , at t_2 , causes e_2 . Both theories postulate the same entities, the only substantive difference is where we locate the causal force, e_1 or s_1^* .

Substance-causation alone is insufficient to capture the moral beliefs embodied in the control principle *et al.*, but together with a robust theory of agent-causation, where agent-substances are complex reason-responsive causes of their own actions, it is maximally consistent with these principles. Agent-substances are the sole, determinate authors of their actions, yet can also cite reasons that played a role in their choices. In light of these unique qualities – qualities that are irreducible to non-agent-substance-causation, agent-causal theories *do* require a more robust ontological commitment than alternative theories (although not as great as the multi-causal “extra factor” theories); however they also explain a great deal more. A world with agent-causation is one where people can be *actually morally responsible* for their actions in the way that we believe they are, it is a world where principles like the control principle are actually applicable to agents like us.

⁷³ See Carl Hoefer (2008).

4.3.1.2 Are Agent-Causes Reason-Responsive?

One troubling aspect of agent-causation is that it is, in a sense, question-begging. We're told agent-causes are intelligent, complex, reason-responsive substances but are given no causal account of how they operate. This has led to continued criticism that agent-causation cannot account for our beliefs about an agent's reason-responsiveness.

For compatibilists, to be reason-responsive is to have reasons play a role in determining one's actions, whether these reasons be desires, compulsions, beliefs, or the like. The libertarian, too, contends that there is usually some relationship between an agent's reasons and what they do, but it cannot be a causal relationship because then the reason is the cause of the agent's actions, rather than the agent herself.

In "Persons and Causes: The Metaphysics of Free Will," Timothy O'Connor purports to offer an account of how an agent can act for reasons without being causally determined by them.⁷⁴ He contends that an agent's choice can be explained in terms of desires the agent has under the following circumstances:

For an agent a desire D is a reason for choice C if and only if:

1. prior to the choice, the agent had an antecedent desire D and believed that action A would satisfy that desire, and
2. the agent's choice C is a choice to engage in action A, and
3. the agent's choice C is caused by the agent in order to satisfy desire D.⁷⁵

The aspect of O'Connor's solution that seems most worthy of noting occurs in 3 above. Here, he contends that a desire, what we're to take as the agent's reason for acting, is a reason the agent acted if and only if the action is done in order to satisfy that desire. For O'Connor, a

⁷⁴ O'Connor, 2000b, page 86.

⁷⁵ Adapted from O'Connor 2000b, page 86, and Clarke 2008.

reason (desire D) is *not* a cause of an action, but rather the agent is the cause and the agent causes it *for* D, not because of D. The difference between an agent doing something *for* D and her being *caused by* D strikes me as an intuitively plausible distinction to make, and a distinction that captures the very essence of the control libertarians are interested in.

One might object that on this view, the reason the agent acts lacks any causal power – and thus the question “Why did she do that?” is *radically different* than the question “Why did it do that?” On one level, I think this is correct – I think that there *is* a fundamental difference between these two sorts of questions. The latter is concerned with a history of causation. However, the latter question *doesn't make sense* when applied to indeterministic events or substances – for example, you might ask “Why did the electron appear there?” (assuming that the place of any given electron in an atom's electron cloud at any given time is undetermined), to which the answer is either (a) for no reason, or (b) because the electron caused itself to be there. *Qua* agent-causal theorist, the latter explanation makes more sense – it was up to the electron, much as an agent's actions are up to the agent. The difference, of course, is in the kind of substance that is doing the causing – that is to say that electrons, for all their wonder, are supposed to be less complex than agent-causes. For the agent-causal theorist, then, one of the aspects that distinguishes an agent from a non-agent substance is the agent's ability to act *for reasons*, as opposed to determined substances who are caused, or undetermined substances that are uncaused.

I think one can *further* object to this view by questioning the intelligibility of acting *for a reason*. On this view, acting for a reason isn't *just* acting *while having* a reason (as some event-causal theorists contend), it is acting with special attention to that reason. The reason in question plays a noncausal role, and yet is supposed to explain the action in question! At first glance, this

may sound bizarre, but it is exactly the sort of role reasons *intuitively* play in free action. For an agent to act for a reason is for that agent to act with that reason occupying a specific epistemological role – one whose role is morally relevant in determining one’s blameworthiness or praiseworthiness for the action in question. There is an epistemological connection between the reason and the agent’s choosing to act that is not there in cases of people who act *with* the same reasons, but not *on* the same reasons. What matters isn’t what the agent does, but what reason or reasons occupy this role while the agent acts. Whether a reason occupies this role is *up to the agent*, but the role is *prima facie* mysterious because it doesn’t explain the choice in a causal role.

The question at hand is what is it for a reason to play a noncausal explanatory role. Instead of a causal explanatory role, one possibility is that reasons for an agent’s actions play a regulative or evaluative role. Here is such an account: (i) For a choice, or action, to be done for a reason is for that reason to occupy a regulative or evaluative epistemological role when the agent acts. (ii) for a reason to occupy this role is such that the choice or action succeeds or fails *for the agent* in terms of whether or not it satisfies this reason. For example, imagine the following case: Joey desires to eat a good meal, and chooses to act to satisfy this desire by cooking a meal. Whether or not Joey succeeds in this action is a question of whether or not his action satisfies his reasons for acting – in this case, whether Joey get to eat a good meal (in the way he thought he would⁷⁶) because of his action. When an action goes well, the agent’s action succeeds in satisfying her reasons – that is to say that her action played a causal role in satisfying

⁷⁶ That is to say that Joey’s action is a success if it satisfies his reason to act (because he wants to eat a good meal) if and only if it satisfies it in the way he intended it to (he cooks a meal intending for it to turn out good, and plans on eating this meal. If he has this intention, then if his cooking of a meal causes his neighbors to bring over a good meal because they know from past experience Joey will fail to cook anything remotely good, and Joey ends up eating this meal instead of the meal he has planned, his cooking a meal does not satisfy his reason for acting, even though it may result in his desire being satisfied in another way).

her reasons. When an action goes poorly (or fails), the agent fails to satisfy her reasons for acting. In this sense, I think, one can maintain a robust semi-causal role of reasons, without reasons playing any causal role in the agent's action.

Furthermore, this account of the role reason plays makes it clear that agents can act on *one set of reasons* without acting on all of the reasons they have to act in one way. For example, a hungry person might go to a business dinner to meet people, rather than to satisfy her hunger.

If reasons play *this role*, the question is how does an agent decide what reasons occupy this special epistemological role? Intuitively this, much like whether they act or not, is up to an agent. The question at hand, then, is whether or not the agent has control over both her reason-selection and her actions that is necessary for moral responsibility. And this question leads us to reconsider the Mind argument.

4.3.1.3 The Mind Argument Revisited

Let's reconsider Peter Van Inwagen's interpretation of the Mind argument – he contends that what doesn't follow deterministically from one's previous states is a result of an indeterministic process, and that no one can determine the outcome of an indeterministic process (2008, 16). If agent-causation is possible, this is false. For the substance-causal theorist, for something to be the result of an indeterministic process *just is* for that something to be caused by a substance such that the outcome is not wholly determined by antecedent facts about the world (coupled with the laws of nature). For an agent-cause to cause something in this way *just is* for it to be up to the agent. Thus, if agent-causation is possible, it is *always the case* that the agent-

substance determines the outcome of the agent's undetermined (by antecedent circumstances) event.^{77 78},

Despite the fact that agent-causation would be sufficient to show that an action is *up to the agent in question*, it's not *prima facie* clear how this grants the agent the sort of control necessary for moral responsibility. If the agent's action is not causally determined by antecedent circumstances, then some argue that it is a matter of chance. And surely no one can be in control (in the way required for moral responsibility) of what is merely chance or luck.

There are, I think, two possible accounts of what this "chance" is. The first account of chance, and the one seemingly advocated by the compatibilist, is that something is a matter of chance if it is not causally determined. However, appeal to this account by critics of incompatibilist theories would be begging the question; it would be an argument by definition, and it is an open question whether or not that definition of "chance" accurately reflects the robust intuitive concept we come to the table with.

A less-question-begging account is that something is a matter of chance if there is no explanation for it's being the case. Critics of agent-causation would be quick to point out that there is nothing in the past that makes it the case that the agent chooses to act as she does, and thus it must be luck that she ends up doing what she does, and luck that so often such agents

⁷⁷ Donald Smith and E.J. Coffman (forthcoming) present a similar argument. They contend that the Mind argument stands or falls on the assumption that an event has only other events among its causes. If this is false, as it would be if agent-causation is true, then the Mind argument is false.

⁷⁸ There is, I think, an uncharitable account of the mind argument that is sometimes proposed, which basically says that no event can be both determined and undetermined, and that indeterministic theories of free will require free events to be both determined and undetermined, and thus no indeterministic account of free will is possible. Ginet (1989, 20) addresses, and refutes this objection. The libertarian doesn't require that an action is both caused by the agent and uncaused by anything; rather the libertarian contends that the agent's action must be caused by the agent, and the agent must not have been caused by antecedent events to act in the way she did. (She can, of course, be prompted (or not) to act, and she can, of course, have her options limited by antecedent circumstances – but what matters is that she, and not the antecedent circumstances, determines what she will do).

would end up acting in accordance with some reasons the necessitarian would point to as causes for a determined agent to act.

The agent causal theorist should be quick to point out that there *is* an explanation for it being the case that an agent acts as she does – because the agent herself willed, or chose, to do so. Furthermore, O'Connor's account of acting for reasons demonstrates why there can be a constant conjunction of reasons with actions without one causing the other.

A second account in which something can be a matter of chance is that it is outside of anyone's control. The result of the coin flip, or the dice roll, or the deal of the cards is a matter of chance because (assuming no card-counting, or it's dice and coin equivalents) although it may be completely nomically necessitated, the outcome is determined by brute facts about the world, and not by the agent's actions. The person flipping the coin may *wish* for heads, but unless she has some atypical ability to control the outcome of the coin flip (say, one garnered through years of practice, and mastery of the environment in which she flips the coin), she has no *say* in its outcome – even though she caused it. This is to say that the coin-flipper may have wanted heads, but unless she somehow manipulated the coin in such a way that she knew would cause the outcome, an outcome of heads, although aimed at, is merely coincidence.

However the difference between the result of a coin flip and the agent's flipping the coin is obvious though – although your average agent can't determine what the outcome is, intuitively she *can* determine that she flips the coin, or pulls the trigger, or raises her arm.

All too often, I think, critics contend that the libertarian agent's actions are a matter of chance *in this way*, but this is just to deny agent-causation and postulate something like an undetermined event that causes the agent to act in such and such a way. This latter account of chance, I think, is the most intuitive, and least question-begging account, but if agent-causation is

possible, then the choices that agent-causes make may not be a matter of chance, rather they could be up to the agent who makes the choice.

4.3.1.4 Is substance-causation incoherent?

Perhaps the most interesting objection to agent-causal theories is that it is based on a mistake. We often use language consistent with substance-causation, but critics suggest that this language is shorthand for a more robust event-causal account that we, in fact, mean. Timothy O'Connor⁷⁹ describes this objection as follows: The critic of agent-causation contends that “We often talk loosely as of inanimate objects as causing certain things to happen” (1998, 377).

For example, we might say something like this: (1a) “The car knocked down the telephone pole.” However, what we mean by this, according to the critic, is that (1b) “The movement of the car, coupled with other aspects of the situation (such as the car’s mass, weight, etc., the position, mass, weight, etc. of the pole, etc.) caused the pole to be knocked down.” Similarly, when we say (2a) “John caused action A,” we *mean* (2b) “John’s reaction to external circumstances was action A.” For the agent-causal theorist, however, when we say (2a), we *mean*, first, (2b’) “John agent-caused action A.” One aspect of (2b’) that critics find unsatisfying is the question of timing. Like the car, John persists through time. However, the car knocks down the pole only at a certain time, and if John causes action A, he does so at a certain time. Event-causation can explain *why* the car knocks down the pole at that time – because the car hit the pole at that time with a certain speed, mass, weight, the pole had a certain position, mass weight, etc... Similarly, event-causation can explain *why* John does action A at the time he does it – because he was caused to by events that he may or may not have played a role in. If it was up to John, and John alone, that he A, and John is an enduring substance, the question is this: Why is it John Aed at the time he did, and not earlier or later? If agent-causation is possible,

⁷⁹ O’Connor, 1998/2001

why does it occur when it occurs, and not at other times? The answer, I think, is simple – when John causes action A, he is reacting to something – the situation that he is presented with. For example, if you ask John why he turned on the TV at 9pm, he might say that he did so to watch the news *which begins at 9pm*.

Some libertarians, like Ginet, believe it's important to secure for the agent the ability to act spontaneously. In such cases, event-causal theorists would need to point to some spontaneous, indeterministic event, perhaps inside the agent's mind, that gets the ball rolling, so to speak. For agent-causal theorists, the whole agent would act spontaneously. All things being equal, the agent-causal theorist's account of spontaneous action is more consistent with not only our commonsense linguistic practices ("John jumped out of his seat, without warning or provocation, and Aed."), but it is also more consistent with principles like the control principle (that John is morally responsible for his Aing).

Regardless on one's stance on spontaneous actions, for *many* actions, agents do, in fact, *react* to external circumstances and events. But there is nothing mysterious about an agent's reacting to a given situation – although, for the agent-causal theorist, it is up to the agent in question what to do in many of these causes, their environment offers a finite number of possible actions. Furthermore, external circumstances often give agents *reasons* to act, and although for agent-causal theorists, reasons do not cause the agent to act, actions can be done for those reasons nonetheless. Thus, an agent can't do action A for reason D unless the agent has reason D, and if an agent doesn't come to have reason D until, say, time $t1$, then before time $t1$, it was not possible for the agent to do action A for reason D at that time.

O'Connor offers a similar account of how agent-causation may work. He contends that although agent-causes may not be able to act *ex nihilo*, or spontaneously, as Ginet claims they

might, they have the ability to govern themselves internally, such that once prompted by external circumstances, the outcome is not wholly causally determined by those circumstances, but is instead up to the agent (O'Connor, 1998, 378).

It seems to me that *spontaneous agent-causation* is no more mysterious in kind than *prompted agent-causation*, as in both cases, what the agent does is said to be up to the agent. The difference is that in spontaneous agent-causation, the agent is in control of not only what the agent does (given the list of alternatives), but *when* the agent does it (given the list of alternatives). All that spontaneous agent-causation requires is a list of alternatives that include different time-stamps – for example, to spontaneously jump out of one's seat and stretch simply requires that one choose from the alternatives to: jump out of one's seat and stretch at time t_1 , jump out of one's seat and stretch at time t_2 , ..., to jump out of one's seat and stretch at time t_n .

There are other ways in which one might call agent-causation incoherent, incomprehensible, and the like. Most, though, focus on the mysterious nature of agent-causation – what is it that makes an agent (or anything, for that matter) an intelligent-cause, a non-arbitrary-cause, and the like? Does it even make sense to say that agents are causes in this way? These questions, I think, are very important.

But there are two ways in which these sorts of questions can be asked – the first, is whether or not this account of causation and action-explanation is consistent with our beliefs and intuitions. The answer, I think, is obvious to everyone but those who have so engrossed themselves in the topic as to render themselves biased – yes! We talk about non-arbitrary, free, undetermined, actions – we talk about agents choosing from amongst alternatives, each of which is open to her at the time of her choosing, and we talk about agents being in charge of their own future, and their own minds. Indeed, when these sorts of claims about someone are in doubt, we

start to wonder whether they're actually a moral agent, whether they're a person, whether they have free will, and can be morally responsible for anything.

The second is whether or not this account of causation and action-explanation is consistent with our ontological beliefs, or beliefs about what exist. This question is traditionally answered by asking whether it is consistent with our scientific beliefs and intuitions, which leads to the question of the next section... is agent-causation compatible with our best scientific theories?

4.3.1.5 Is agent-causation compatible with our “best scientific theories”?

In Derk Pereboom's “Living Without Free Will,” he contends, “Agent causation is a coherent possibility, but it is not credible given our best physical theories” (2002, 477).⁸⁰ But if not agent-causation, what is the alternative? Event-causation is the most cited alternative, but this seems to fare no better. Carl Hoefer points out that “... neither philosophers' nor laymen's

⁸⁰ In *Living Without Free Will* Pereboom, he goes into detail about how agent-causal theories of free will are said to be inconsistent with our scientific views (2001, 69-88). Although Pereboom argues agent causation is coherent, he contends that we lack it because it is incompatible with our best physical theories. While his arguments to this effect are too many to cover effectively in this work, it strikes me that Pereboom employs a strategy found in elsewhere (for example Flanagan, 2002) that is, I think, turn on an uncharitable account of the causation agent-causal theorists advocate. He asks, “Given our scientific understanding of the world, how could there be agent-caused decisions that are freely willed in the sense required for moral responsibility?” He then claims “if agent-causes are to be capable of such free decisions, they would require the power to produce deviations from the physical laws – deviations from what these laws would predict and from what we would expect given these laws” (Pereboom, 2001, 79). By “physical laws”, Pereboom means the laws of physics. The contention is that an agent-cause would require one to break the laws of physics – but this, I think, even the agent-causal theorist would find implausible. What strikes me as odd about this argument is that no one (seriously) proposes that agent-causation breaks the laws of physics. Rather agent-causal theorists contend that agent-causation can be explained in terms of the laws of physics. It is the case that our current best scientific theories fail to explain the unique account of causation required for moral responsibility by the agent-causal theorist, but they fail to offer a robust, intuitive account of anything like an event either. But surely this omission from the current state of the art physical laws doesn't mean that we ought to give up all philosophically relevant accounts of an event. It may very well be the case that our best physical theories are *prima facie* inconsistent with agent-causation, I think detractors like Pereboom unduly help themselves to intuitive plausibility when they contend that the agent-causal theorist advocates a type of causation that violates the physical laws. This is false. Rather, what they advocate is a type of causation that is *prima facie* inconsistent with our current best physical laws, but also contend that our best physical laws are, at best, incomplete; at worst, full of holes and mistakes. Pereboom paints a picture of the agent-causal theorist trying to force a puzzle piece into a near-completed puzzle, while in reality it's more like the agent-causal theorist is trying to do her best to fit a puzzle piece into a puzzle that is missing half of it's pieces – a puzzle piece that came with the box, and looks as if it belongs to the puzzle. We should balk at the agent-causal theorist if she contends that our physical laws are routinely broken by agent-causes, but her position is *far* less preposterous if she contends that the proper account of our physical laws will explain agent-causation, much as they explain non-agent-causation.

conceptions of *events* have any correlate in any modern physical theory....” (2008). In an aside, he explains what may have lead so many philosophers astray, “Some philosophers are misled on this point by the fact that some now-defunct presentations of Special Relativity theory seem to be grounded on an ontology of events. But Special Relativity does not need to be so presented, nor were the “events” used anything like common sense events.” Perhaps we ought to go back to Hume’s account of causation, as a constant conjunction of cause and effect⁸¹ – but this account seems perfectly consistent with either event-causation or substance-causation, and – worse still – gives us no account of the science underlying causation.

Adding to the confusion, many event-causal theorists claim agent-causation is reducible to event-causation, in manners similar to those O’Connor cited (discussed above). However, I see no reason the substance-causal theorist couldn’t offer similar claims about the reduction of events to substances, including in some cases agent-substances. The only difference, it seems, between the two projects is that agent-causation is able to capture this concept of an action’s arising from the agent, *qua* enduring substance, while event-causation is stuck positing either a fickle, disappearing instance of indeterminacy or a lifetime of determinism. Indeed, agent-causation affords us all of the causal control as deterministic theories of free will (insofar as the agent is the *cause* of her actions), while securing the ability to do otherwise (required by AP) in the agent herself, rather than in some indeterministic event.

After advancing an agent-causal view, in “Human Freedom and the Self,” Roderick Chisolm says the following:

If we are responsible, and if what I have been trying to say is true, then we have a prerogative which some would attribute only to God: each of us, when we act, is a prime

⁸¹ David Hume *A Treatise on Human Nature* 2.3.2.4.

mover unmoved. In doing what we do, we cause certain events to happen, and nothing — or no one — causes us to cause those events to happen (Chisholm, 1964/2002, 55–56).

When acting as an agent-cause, Chisholm contends, we have an ability found nowhere else in nature. He follows up this by claiming:

This means that, in one very strict sense of the terms, there can be no science of man. If we think of science as a matter of finding out what laws happen to hold, and if the statement of a law tells us what kinds of events are caused by what other kinds of events, then there will be human actions which we cannot explain by subsuming them under any laws. We cannot say, ‘It is causally necessary that, given such and such desires and beliefs, and being subject to such and such stimuli, the agent will do so and so.’ For at times the agent, if he chooses, may rise above his desires and do something else instead (Chisholm, 1964/2002, 56).⁸²

⁸² In the 2002 book *The Problem of the Soul* Owen Flanagan takes issue with Chisholm’s conclusion, claiming “The scientific image says that we are animals that evolved according to the principles of natural selection. Although we are extraordinary animals we possess no capacity that permits us to circumvent the laws of cause and effect (Flanagan, ix).” Flanagan, I think, misses the point – for Chisholm’s claim doesn’t amount to the claim that agents circumvent cause and effect – rather, it amounts to the claim that there are some causal processes that are not determined by antecedent circumstances, and are instead determined by agents. Such an ability would surely be valuable in at least one sense – insofar as agent-causes can choose from multiple options, there is a chance that they may choose the right options instead of the wrong option. Thus, in a situation where another animal is causally determined to go left (say, because it looks the most appealing), the agent-cause can cause herself to go left or right. If left is a trap, having the ability to go against the best reason and go right is an ability worth having. Still, Flanagan’s point may be that, on average, having this ability is a detriment. Which animal would more often pass the test of natural selection – the animal that always does what it has the best reasons to do, or the animal that only sometimes does what it has the best reasons to do (because it has the ability to do otherwise)? If we assume both animals have relatively correct world-views, surely the former. Thus, Flanagan’s point may be that if evolution is true, natural selection would have favored strict causally-deterministic action-causation over libertarian indeterministic action-causation. This objection is equally suited against both agent-causation and event-causation. The problem with this sort of objection is threefold – first, it seems to assume that natural selection doesn’t make mistakes. It is *prima facie* plausible natural selection can select an animal that isn’t the best and brightest, but merely the only one left (or the one that sows the most seeds, as it were). Second, natural selection often selects animals for traits that are valuable at a time, but may be less valuable (but, say, not harmful) at other times. For example, consider the gene that causes hemochromatosis, a condition that causes iron to accumulate in a human being’s organs, eventually leading to death. Once thought to be rare, recent research has shown the gene to be common. One explanation for the frequency of the gene is that it cuts off the iron available to iron-feeding bacteria like the bubonic plague. Hemochromatosis is, at best, contingently valuable, but without the prevalence of such diseases, having the gene for hemochromatosis is, all things being equal, detrimental. Third, it is not at all clear that having the God-like ability to be one’s own cause is at all *less* valuable from an evolutionary standpoint than the

If Chisholm is right, then *of course* agent-causation is inconsistent with our best scientific theories, as our scientific theories are concerned with regular laws, and mankind transcends this ability. Despite this, it is a mistake to say that the workings of free agents is outside the realm of science – it seems to me that the end of science isn't just a matter of explaining natural laws, rather it has a more lofty goal: to explain *everything* – even those things that cannot be explained in terms of causal laws. In light of this, were the agent-causal views *really* inconsistent with the our best scientific theories, it would count as evidence against the the theory that agent-causation may be true at the actual world. However, philosophers who offer a detailed metaphysics of causation tend to subscribe to either agent-causation or event-causation, and neither one seems to *obviously* more consistent with our best scientific theories than the other. In light of this, the question becomes which theory of causation better explains our moral beliefs and intuitions, as they are on roughly equivalent scientific footing. If the answer to this question is agent-causation, as I've argued it is, then we have every reason to believe that agent-causation is possible, even likely, at the actual world.⁸³ Since we have good reason to think agent-causation is possible at the actual world, we have good reason to think that libertarian moral responsibility is possible at the actual world.

alternative. The ability to pick from multiple alternatives is *prima facie* more valuable than being caused to act in one way alone. For every horror story of some agent, acting freely, doing something dangerous and stupid, there is an innovation story of some agent, acting freely, doing something others had never thought of before, doing things in a new way. Surely, in this sense, the ability to be one's own cause offers great risk and great reward. To the extent that agent-causes not only control their actions, but are non-arbitrary, intelligent causes, the evolutionary benefit of the latter surely outweighs the evolutionary drawback of the former.

⁸³ Especially in virtue of the approach I advocate in chapter 1.

Chapter 5: Conclusion

5.1 This has been a defense of libertarianism

According to libertarianism, to be morally responsible for something requires that you own that thing; if moral agents are morally responsible for anything, they are so for their own actions because they are the free causes of those actions. Libertarianism is maximally consistent with our commonsense beliefs about moral responsibility; the vast majority of our deeply held moral principles only make sense on the assumption that what the moral agent does is up to her, that she is the free cause of her actions. Only then does it make sense to blame or praise *her*, rather than some external cause, such as genetics, contingent events, or even some indeterminism at the subatomic level.

Libertarianism is maximally consistent with the control principle and the principle of alternate possibilities, and in virtue of chapters 2 and 3, I believe we have good reason to think that these principles adequately capture the core of our commonsense moral beliefs and intuitions. In recent years, these principles have been challenged. The problem of moral luck threatened to undermine our confidence in the control principle, much as Frankfurt's famous case threatened to undermine the near-universal agreement about the truth of principle of alternate possibilities. In chapter 2 I argued that libertarianism is immune to resultant luck, situational luck, and circumstantial luck, and in chapter 4 I show that libertarianism, despite requiring indeterminism, is immune to causal luck because it requires agent-causation, or something very much like it. In chapter 3 I argued that Frankfurt's argument against the principle of alternate possibilities failed because his supposed counterexample fails to demonstrate the falsity of any charitable interpretation of the principle. At best, Frankfurt has shown that moral agents do not require alternate outcomes to be morally responsible, but to hold

agents morally responsible for the results of their actions, rather than their free choices that caused their actions, would be to introduce resultant luck into the picture.

Libertarianism requires indeterminism, and in chapter 4 I argued that indeterministic theories of moral responsibility only make sense if they appeal to agent-causation, or something very much like it. I've shown how agent-causal indeterministic theories can answer three criticisms that event-causal indeterministic theories cannot. At best these arguments show that moral agents lack control, or authorship, over their actions if event theories of causation are true. If events cause other events, and moral agents aren't events, but substances, then moral agents can never be the author of their own actions in the way libertarianism requires for moral responsibility. However, if substance-causation, in particular agent-causation, is possible, then substances, such as agents, can cause events. These agent-causes can thus be morally responsible for the events they have complete control over, their own choices and intentions. Following this, I defended the theory of substance causation from five traditional criticisms. I argued that substance-causation is no more ontologically complex than event-causation, that substance-causation is no less *prima facie* compatible with our best scientific theories, and that we have good reason to believe that we may be agent-causes of the kind required by libertarianism – because only then can we explain our belief in principles like the control principle and principle of alternate possibilities in a way that does not introduce massive error into our belief sets.

5.2 What next for the critic of libertarianism?

What next, then, for the critic of libertarianism? In this section, I will discuss three directions the critic of the libertarian should explore. First, it seems to me that *record luck*, while not *moral luck*, may cause problems for libertarians. Second, I propose what I believe to be the best

possible avenue Frankfurt-theorists can pursue in an effort to find a true Frankfurt case, one where the agent in question is nontrivially morally responsible despite lacking alternate possibilities. Third, I discuss the situations under which event-causal theorists may have good reason to question the legitimacy of substance-causation.

5.2.1 Record Luck

Many philosophers who argue that moral luck exists and undermines the control principle no doubt find my treatment of moral luck in chapter 2 infuriating. I argue moral luck exists if and only if an entry in one's moral record is determined, even in part, by luck, or circumstances outside of their control. The number of entries in one's moral record, and whether or not one is a moral agent (rather than a non-agent such as a chair) are *not* instances of moral luck; rather they are instances of what I call *record luck*. Record luck exists. However, it is my position that no one is properly any more or less morally praiseworthy or blameworthy in virtue of record luck. There are two types of record luck – (1) luck in one's having a moral record, and (2) luck in the number of entries in one's moral record. It seems to me that the critic of the control principle (or libertarianism in general) might argue that record luck of either of these types would undermine our moral intuitions and thus collapse the web of beliefs that lead many of us to believe libertarianism is true.

In regards to (1), suppose a morally blameworthy agent were to contend that it was her bad fortune to be a person, rather than a chair, because as a non-person she would be neither praiseworthy or blameworthy, and surely this is morally preferable to her current state of being morally blameworthy. I suppose the argument would go something like this:

1. It is only luck that I am a person, and not a non-person.
2. Only persons can be morally blameworthy

3. Thus, it is a matter of luck that I am morally blameworthy.

This argument is invalid; the conclusion should be:

3'. Thus, it is a matter of luck that I can be morally blameworthy.

While for the libertarian it is a matter of luck whether or not you *can* be blameworthy, it is not a matter of luck whether you *are* morally blameworthy – that depends upon you. That *just is* what it is to be a moral agent, a person.

I take it that record luck of type (2) is typically such a contentious subject because it seems inherently unfair that (a) some moral agents get more chances than others, and that (b) some get chances to do really great, or horrible, things that others never do. In chapter 2, I argue that the idea of (b) is a mistake – I contend that it is at least *prima facie* plausible that moral agents can choose to do absolutely horrible things on their own. For example, in chapter 2 I suggest that Janine, who opens the door to find girl scouts selling cookies, might form the intention to murder someone the next time she believes she can get away with it. It is *prima facie* plausible that sufficiently developed moral agents can form intentions of this kind in most situations where they can act *qua* moral agents, and that forming such an intention *just is* choosing to act on that intention should the right conditions arise. In other words, it doesn't make sense to say that some moral agents turn out more or less praiseworthy or blameworthy because of the quality of the moral tests they are confronted with, because moral tests are concerned with one's intentions and choices, not the results of the situations they face.

However, it seems undeniable that some moral agents live longer than others, and have more opportunities to act than others. In chapter 2 I discuss a case where Chris and Kris are morally identical agents up until a point when Kris dies at time *t*. Up until time *t*, their moral records would be identical, however after *t*, we're supposed to believe that their moral records

would be different solely because Kris died and Chris lived. The idea is that the final summary of Kris's life is different than the final summary of Chris's life because the latter lived longer, and did more things, than the former. This seems to me to be a mistake.

Whether or not Kris and Chris's moral records necessarily differ because of Kris's death seems to lead us into dangerous territory – either there is an afterlife or there is not. Only moral agents can be morally praiseworthy or morally blameworthy; thus if Kris no longer exists, or is no longer a moral agent, after she dies, then it doesn't make sense to discuss her moral record at all. A summary of her moral worth at her death may be comforting to those who knew her, but it doesn't help us treat *her* as she deserves to be treated, because she no longer exists.

Suppose, though, that Kris still exists as a moral agent after her death, as an immortal soul or something along these lines. It seems to me that Kris can continue to make choices that are morally equivalent to Chris's even after death. If all moral agents are immortal in this sense, then it seems to me that – again – one's total moral worth in life is arbitrarily meaningless – at time t_2 , Chris and Kris may still be morally equivalent despite existing in different forms. There would be no such thing as one's total moral worth, rather only cursory summaries of one's moral worth at a given time.

Either moral agents have finite existences or infinite existences. In either case, this has no bearing on what they deserve – whether they are praiseworthy or blameworthy – at any of point along the way. If moral agents have finite existences, there is no point to calculating their total moral worth as they no longer exist to blame or praise (their moral record ceases with them); if they have infinite existences, there can be no total moral worth, only cursory summaries of one's moral worth until that time.

Some people might believe that the passing from life to an afterlife might mark an important period where the agent is judged by a more accurate moral judge (whether God, or natural forces) than the agent can be judged by during life. For people who maintain such a belief, then record luck may appear to be a problem. However, the accuracy of a third party moral judge at any point in one's life, and the following (accurate or inaccurate) punishment or praise, is neither a matter of moral luck nor record luck. It is luck in how we are treated. But how we are treated, and whether we are treated as we deserve, morally, is ultimately a matter of luck. Some moral luck theorists might contend that this undermines the control principle because, ultimately, we lack complete control over how others judge us, such that moral actions might not always bring actual praise, and immoral actions might not always bring actual blame. However, this is just another case of resultant luck that doesn't affect one's actual moral record, and thus isn't a case of moral luck.

5.2.2 Frankfurt Cases

I hold Harry Frankfurt's approach in *Alternate Possibilities and Moral Responsibility* in high regard; to demonstrate the falsity of the principle of alternate possibilities, he attempts to construct a situation where an agent lacks alternate possibilities, yet our commonsense moral intuitions tell us he is undeniably morally responsible. The problem is that his example falls well short of this goal – he shows that moral responsibility doesn't require alternate outcomes, but fails to show that it doesn't require alternate possibilities.

The problem with Frankfurt's approach, and one that has been repeated in the various attempts at constructing such a case that have followed, is that the case is designed to appear to the third party observer as if everything was normal, but then the author stipulates some hidden aspect of the case designed to cut off alternate possibilities. The Frankfurt case advocate *seems*

to take advantage of our ignorance about how the mind works, and assumes that given this new information about how the mind works, whether atypical nor not, we'd be inclined to make the same moral judgment that we would make were we ignorant of these hidden aspects. But this is absurd. In those Frankfurt cases where the agent clearly lacks alternate possibilities, the mechanisms that cut off these alternate possibilities are *prima facie* agency-undermining, and thus good reason to revise our intuitions about the agent's moral responsibility.

In a sense, the typical Frankfurt case is deceptive – we're given a situation where we'd typically believe the agent has alternate possibilities and as such is morally responsible for what she's done, and then – *surprise* – told the agent lacks alternate possibilities, with the implication being that we've already committed to saying that the agent is morally responsible and should stick to these judgments despite the unintuitive aspects of the case. This approach is fundamentally flawed.

Those who believe Frankfurt was on the right track need to revise how they approach constructing a Frankfurt case. The Frankfurt case supporter ought to attempt to construct a case where the agent's lack of alternate possibilities is obvious even to the third party observer, yet where it is equally obvious that the agent is morally responsible for what she has done *even though* she lacks alternate possibilities. I do not believe such a case exists, *however* this approach is the *only* approach that can satisfy Frankfurt's goal of constructing a case where *everyone*, whether libertarian incompatibilist or necessitarian compatibilist, would agree the agent is morally responsible despite lacking alternate possibilities. If the necessitarian is right, *every* case where an agent is morally responsible is necessarily a true Frankfurt case, but Frankfurt's goal was to convince the compatibilist and incompatibilist alike.

5.2.3 On Agent-causation

One option open to the critic of agent-causation, or substance-causation in general, is to come up with an alternative theory that doesn't call into question a large majority of our commonsense moral beliefs in the way that event-causal theories do. Short of this, the critic has two options – (1) show that agent-causation, or substance-causation, is incoherent, or (2) show that the moral beliefs are consistent with agent-causation are somehow flawed – either (a) they are not as widespread as the libertarian claims, or (b) they are the result of a rather obvious, in retrospect, series of flaws in our belief-forming processes. Regarding (1), the concept of substance-causation is *prima facie* neither more ontologically complex or incoherent than it's alternatives, most notably event-causation. Agent-causation is often attacked on the basis that it is an uncaused, but purposeful kind of causation, but this just confuses the issue. Agent-causation is not uncaused, it is caused – by the agent herself – and if event-causation can, in any way, be said to be purposeful, agent-causation can surely be said to be purposeful in the same manner.

(2)(b) is a rather interesting approach, although far too often critics of (2)(b) presuppose universal causal determinism, and thus beg the question. *Of course* if universal causal determinism is true, a large portion of our commonsense moral beliefs, as well as commonsense beliefs about the world in general, turn out to be false, and by assumption these beliefs would have been causally determined by various belief-forming processes that fail to accurately represent the world. The challenge for the (2)(b) advocate is to advance a theory to explain these beliefs that does not presuppose the very theory that it is attempting to make more palatable.

Perhaps the best way to approach (2)(a) is to engage in experimental philosophy, along the lines of what Shaun Nichols and Joshua Knobe have done recently.⁸⁴ Nichols and Knobe

84 See Nichols, Shaun and Knobe, Joshua, 2007, "Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions," *Nous* 41, 663-685.

conducted a series of experiments that seem to show that people have both compatibilist and incompatibilist intuitions about moral responsibility. Subjects were presented with one of two near- identical cases where the case's antagonist causes harm to others, but has no control over their action. In one variation, the harm is described in detail, in the other the harm is described dispassionately. Nichols and Knobe contend that the responses consistent with compatibilism may have been a byproduct of affective-performance error; the cases which tended to produce compatibilist-style responses from subjects tended to have language which would agitate the subject's emotions. Nichols and Knobe contend that this increased emotional agitation may lead to errors in judgements of moral responsibility.

Perhaps the largest problem with experimental philosophy is that the very distinctions that philosophers are interested in are often hard to capture through the experimental process. When subjects are presented with one case there is nothing to compare it to, when they are presented with multiple cases there is too much temptation to distinguish between the two. The very act of presenting laymen with the robust, philosophically distinct cases necessary to gauge whether they make certain distinctions might cause the agents to make the distinctions they wouldn't normally make. Still, to the extent that agent-causal libertarianism appeals to its consistency with commonsense moral beliefs, any attempt to gauge whether or not these beliefs are actually commonsense is *prima facie* relevant.

5.3 What next?

According to the libertarian, moral agents are morally responsible for their free choices and intentions, not the results of their actions. The libertarian, it seems, has two distinct projects to pursue – first, insofar as our justice system is concerned with treating people as they ought to be treated, the government has a purely practical interest in developing a technology that would

allow us to access the otherwise private minds of others, specifically those whom we have good reason to believe acted immorally, such as in virtue of the results of their actions. If technology progresses at anywhere near the rate it has during my lifetime, I think we have good reason to believe this technology is not only possible, but likely within several decades. Such technology will no doubt raise many important moral questions, and should not be used in the everyday lives of people. However, to the extent that justice systems are concerned with discovering the truth of how people ought to be treated, such a tool would allow legal systems to better conform with our commonsense moral beliefs and, ultimately, allow us to treat people how they deserve to be treated more often.

Second, because what determines one's moral responsibility is one's inherently private mental states, including the epistemological connections between one's reasons and their choices and intentions, moral philosophers should develop robust theories concerning what counts as morally praiseworthy and/or morally blameworthy reasons to act, as well as investigate the moral character of negligence and willful ignorance, and the role these traits play in determining one's moral praiseworthiness and/or blameworthiness. Ethics is concerned with answering the question "What is the right thing to do?", and often times the right thing to do is to treat people how they deserve to be treated – to praise the praiseworthy, and blame the blameworthy. This includes blaming the person who did the right thing for the wrong reasons, and praising the person whose good intentions had bad unintended consequences. Some people say that the road to hell is paved with good intentions, but surely this road is at least more a circuitous route to hell than the road paved with bad intentions.

Bibliography

- Adams, Robert Merrihew, 1973, "Middle Knowledge", *The Journal of Philosophy*, 70: 552-554.
- Adams, Robert Merrihew, 1977, "Middle Knowledge and the Problem of Evil", *American Philosophical Quarterly*, xiv, 2: 109-17
- Adams, Robert Merrihew, 1985, "Involuntary Sins", *The Philosophical Review*, 94: 3-31.
- Adler, Jonathan, 1987, "Luckless Desert is Different Desert", *Mind*, 96: 247-249.
- Almeida, M., and Bernstein, M., 2003, "Lucky Libertarianism," *Philosophical Studies*, 113: 93–119.
- Anderson, Elizabeth S., 1999, "What is the Point of Equality", *Ethics*, 109: 339-349.
- Andre, Judith, 1983, "Nagel, Williams, and Moral Luck", *Analysis*, 43: 202-207
- Aristotle, 2001, *Nicomachean Ethics* in *The Basic Works of Aristotle* ed. Richard McKeon. Randomhouse
- Arneson, Richard J. and Hurley, Susan, 2001, "Luck and Equality" *Proceedings of the Aristotelian Society*, Supplementary Volumes, Vol. 75: 51-90
- Arneson, Richard J., 2002, "The Smart Theory of Moral Responsibility and Desert", forthcoming <http://philosophy.ucsd.edu/faculty/rarneson/smarttheory2.pdf>
- Ayer, A. J., 1954, "Freedom and Necessity," In his *Philosophical Essays*. New York: St. Martin's Press: 3-20
- Balaguer, Mark, 1999, "Libertarianism as a Scientifically Reputable View," *Philosophical Studies*, 93: 189–211.
- Balaguer, Mark, 2004, "A Coherent, Naturalistic, and Plausible Formulation of Libertarian Free Will," *Noûs*, 38: 379–406.
- Berkeley, George, 1710/1998, *A Treatise Concerning the Principles of Human Knowledge*, ed. Jonathan Dancy. Oxford: Oxford University Press.
- Blumenfeld, David, 1971, "The Principle of Alternate Possibilities," *The Journal of Philosophy*, Vol. 68, No. 11: 339-345.
- Browne, Brynmor, 1992, "A Solution to the Problem of Moral Luck", *The Philosophical Quarterly*, 42: 343-356.
- Chisolm, Roderick M., 1964//2002, "Human Freedom and the Self", The Lindley Lecture, University of Kansas; reprinted in Kane 2002
- Chisholm, Roderick M., 1966, "Freedom and Action." *Freedom and Determinism*, ed. Keith Lehrer, 11-44. New York: Random House.
- Chisholm, Roderick M., 1971, "Reflections on Human Agency." *Idealistic Studies* 1: 33-46.
- Chisholm, Roderick M., 1976a, "The Agent as Cause." *In Action Theory*, ed. Myles Brand and Douglas Walton, 199-211. Dordrecht: D. Reidel.
- Chisholm, Roderick M., 1976b, *Person and Object: A Metaphysical Study*. La Salle: Open Court.
- Chisholm, Roderick M., 1978, "Comments and Replies," *Philosophia*, 7: 597–636.
- Clarke, Randolph, 1993, "Toward a Credible Agent-Causal Account of Free Will." *Nous* 27: 191-203.
- Clarke, Randolph, 1996, "Agent Causation and Event Causal in the Production of Free Action." *Philosophical Topics* 24, no. 2: 19-48.
- Clarke, Randolph, 2008, "Incompatibilist (Nondeterministic) Theories of Free Will," *Stanford Encyclopedia of Philosophy*, September 12, 2008. <http://www.seop.leeds.ac.uk/entries/incompatibilism-theories/>
- Cohen, G.A. 1989, "On the Currency of Egalitarian Justice", *Ethics*, 99: 906-44.

- Copp, David, 1997, "Defending the Principle of Alternate Possibilities: Blameworthiness and Moral Responsibility," *Nous*, xxxi, 4: 441-56.
- Curren, Randall, 1989, "The Contribution of Nicomachean Ethics iii.5 to Aristotle's Theory of Responsibility," *History of Philosophy Quarterly* 6: 261-277.
- Curren, Randall, 2000, *Aristotle on the Necessity of Public Education*, New York: Roman & Littlefield.
- Davis, M., 1986, "Why Attempts Deserve Less Punishment Than Completed Crimes", *Law and Philosophy*, 5: 1-32.
- Dennett, Daniel C, 1973, "Mechanism and Responsibility," In Lehrer (1973).
- Dennett, Daniel C, 1978, "On Giving Libertarians What They Say They Want." *Brainstorms: Philosophical Essays on Mind and Psychology*, 286-99. Montgomery, Vt. : Bradford Books.
- Dennett, Daniel C, 1984, *Elbow Room: The Variety of Free Will Worth Wanting*, Cambridge, Massachusetts: Bradford Books.
- Domskey, Darren, 2004, "There is No Door: Finally Solving the Problem of Moral Luck," *The Journal of Philosophy* 101: 445-464.
- Donagan, Alan, 1987, *Choice: The Essential Element in Human Action*. London: Routledge & Kegan Paul.
- Eccles, John, 1970, *Facing Reality*, New York: Springer-Verlag.
- Ekstrom, Laura Waddell, 2000, *Free Will: A Philosophical Study*. Boulder: Westview Press.
- Ekstrom, Laura Waddell, 2003, "Free Will, Chance, and Mystery." *Philosophical Studies* 113: 153-80.
- Enoch, D. and A. Marmor, 2007, "The Case Against Moral Luck," *Law and Philosophy* 26: 405-436.
- Feinberg, Joel, 1970, *Doing and Deserving: Essays in the Theory of Responsibility*, Princeton, NJ: Princeton University Press.
- Feinberg, Joel, 1995, "Equal Punishment for Failed Attempts: Some Bad But Instructive Arguments Against It", *The Arizona Law Review*, 37: 117-133.
- Fischer, John Martin, 1995, "Libertarianism and Avoidability: A Reply to Widerker." *Faith and Philosophy* 12: 119-25.
- Fischer, John Martin and Ravizza, Mark, 1998, "Morally Responsible People without a Freedom," from *Responsibility and Control: A Theory of Moral Responsibility*. New York: Cambridge, from *The Determinism and Freedom Philosophy Website*, accessed: 7-05-09
<http://www.ucl.ac.uk/~uctytho/dfwCompatFischerRavizza.htm>
- Fischer, John Martin; Kane, Robert; Pereboom, Derk; and Vargas, Manuel, 2007, *Four Views on Free Will*, Oxford: Blackwell.
- Flanagan, Owen, 2002, *The Problem of the Soul*, New York: Basic Books.
- Frankfurt, Harry G., 1969, "Alternate Possibilities and Moral Responsibility", *The Journal of Philosophy*, Vol. 66, No. 23: 829-839.
- Frankfurt, Harry G., 1971, "Freedom of the Will and the Concept of a Person," *Journal of Philosophy* 68: 5-20.
- Gaskin, R., 1993, "Conditionals of Freedom and Middle Knowledge," *The Philosophical Quarterly*, Vol. 43, No. 173, Special Issue: Philosophers and Philosophies, pp. 412-430.
- Ginet, Carl, 1989, "Reasons Explanations of Action: An Incompatibilist Account," *Philosophical Perspectives*, 3: 17-46.

- Ginet, Carl, 1990, *On Action*, Cambridge: Cambridge University Press.
- Ginet, Carl, 1997, "Freedom, Responsibility, and Agency," *Journal of Ethics*, 1: 85–98.
- Ginet, Carl, 2002, "Reasons Explanations of Action: Causalist versus Noncausalist Accounts," In *The Oxford Handbook of Free Will*, ed. Robert Kane, 386–405. New York: Oxford University Press.
- Ginet, Carl, 2007, "An Action Can Be Both Uncaused and Up to the Agent," In *Intentionality, Deliberation and Autonomy: The Action-Theoretic Basis of Practical Philosophy*, eds. Christoph Lumer and Sandro Nannini, 243–55. Aldershot, UK: Ashgate.
- Goetz, Stewart, 1988, "A Noncausal Theory of Agency," *Philosophy and Phenomenological Research*, 49: 303–16.
- Goetz, Stewart, 1997, "Libertarian Choice," *Faith and Philosophy*, 14: 195–211.
- Goetz, Stewart, 2000, "Naturalism and Libertarian Agency," In *Naturalism: A Critical Analysis*, ed. William Lane Craig and J. P. Moreland, 156–86. London: Routledge.
- Goodman, Nelson., 1955, *Fact, Fiction, and Forecast*, Cambridge, MA: Harvard University Press.
- Glover, Jonathan, 1970, *Responsibility*. New York: Humanities.
- Greco, John, 1995, "A Second Paradox Concerning Responsibility and Luck", *Metaphilosophy*, 26: 81-96.
- Griffith, Meghan, 2005, "Does Free Will Remain a Mystery? A Response to van Inwagen," *Philosophical Studies*, 124: 261–69.
- Griffith, Meghan, 2007, "Freedom and Trying: Understanding Agent-Causal Exertions," *Acta Analytica*, 22: 16–28.
- Haji, Ishtiyaque, 1999a, "Indeterminism and Frankfurt-type Examples," *Philosophical Explorations*, 2: 42–58.
- Haji, Ishtiyaque, 1999b, "Moral Anchors and Control," *Canadian Journal of Philosophy*, 29: 175–204.
- Haji, Ishtiyaque, 2000a, "Indeterminism, Explanation, and Luck," *Journal of Ethics*, 4: 211–35.
- Haji, Ishtiyaque, 2000b, "Libertarianism and the Luck Objection," *Journal of Ethics*, 4: 329–37.
- Haji, Ishtiyaque, 2000c, "On the Value of Ultimate Responsibility," In *Moral Responsibility and Ontology*, ed. Ton van den Beld, 155–70. Dordrecht: Kluwer.
- Haji, Ishtiyaque, 2001, "Control Conundrums: Modest Libertarianism, Responsibility, and Explanation," *Pacific Philosophical Quarterly*, 82: 178–200.
- Hansen, Jorgen, 2009, *Do We Really Have Control? New Problems Facing Libertarian Free Will*, 2009 Intermountain West Student Philosophy Conference, University of Utah, Department of Philosophy, March 26-29.
<http://philpapers.org/profile/1647>
- Hobart, R. E., 1934, "Free Will as Involving Indeterminism and Inconceivable Without It," *Mind* 43: 1-27.
- Hoffman, Joshua and Rosenkrantz, Gary, 2002, *Exploring the Divine Attributes*, Malden, MA.
- Hoefer, Carl, 2008, "Causal Determinism," *Stanford Encyclopedia of Philosophy*, April 1, 2008
<http://plato.stanford.edu/entries/determinism-causal/>
- Hobbes, Thomas, 1997. *Leviathan*, R.E. Flatman & D. Johnston, eds. New York: W.W. Norton & Co.
- Hume, David, 1975, *A Treatise of Human Nature*, edited by L. A. Selby-Bigge, 2nd ed. revised by P.H. Nidditch, Oxford: Clarendon Press.

- Hume, David, 1975, *Enquiry concerning the Principles of Morals*, edited by L. A. Selby-Bigge, 3rd edition revised by P. H. Nidditch, Oxford: Clarendon Press.
- Hume, David, 1999, *An Enquiry concerning Human Understanding*, edited by Tom L. Beauchamp, Oxford: Clarendon Press.
- Hurley, Susan, 2001, "Luck and Equality", *Proceedings of the Aristotelian Society*, supplementary volume, 75: 51-72.
- Hurley, Susan, 2003, *Justice, Luck, and Knowledge*, Cambridge: Harvard University Press.
- Jensen, Henning, 1984, "Morality and Luck", *Philosophy*, 59: 323-330
- Latus, Andrew, 2000, "Moral and Epistemic Luck", *Journal of Philosophical Research*, 25: 149-172.
- Latus, Andrew, 2001, "Moral Luck," *The Internet Encyclopedia of Philosophy*, ed. J. Feiser . <http://www.utm.edu/research/iep/m/moralluc.htm>
- Levi, D., 1989, "What's Luck Got to Do With It?", *Philosophical Investigations*, 12: 1-13
- Lewis, David, 1989, "The Punishment That Leaves Something to Chance", *Philosophy and Public Affairs*, 96: 227-242.
- Kalderon, Mark, 2005, *Moral Fictionalism*, Oxford: Clarendon Press.
- Kane, Robert, 1985, *Free Will and Values*. Albany: State University of New York Press.
- Kane, Robert, 1989, "Two Kinds of Incompatibilism." *Philosophy and Phenomenological Research* 50: 219-54.
- Kane, Robert, 1994, "Free Will: The Elusive Ideal." *Philosophical Studies* 75: 25-60.
- Kane, Robert, 1996a, "Freedom, Responsibility, and Will-Setting." *Philosophical Topics* 24, no. 2: 67-90.
- Kane, Robert, 1996b, *The Significance of Free Will*. New York: Oxford University Press.
- Kane, Robert, 1999a, "On Free Will, Responsibility and Indeterminism." *Philosophical Explorations* 2: 105-21.
- Kane, Robert, 1999b, "Responsibility, Luck, and Chance: Reflections on Free Will and Indeterminism," *Journal of Philosophy* 96: 217-40.
- Kane, Robert, 2000a, "The Dual Regress of Free Will and the Role of Alternative Possibilities." *Philosophical Perspectives* 14: 57-79.
- Kane, Robert, 2000b, "Free Will and Responsibility: Ancient Dispute, New Themes." *Journal of Ethics* 4: 315-22.
- Kane, Robert, 2000c, "Responses to Bernard Berofsky, John Martin Fischer and Galen Strawson," *Philosophy and Phenomenological Research* 60: 157-67.
- Kane, Robert, 2002a, "Some Neglected Pathways in the Free Will Labyrinth," *The Oxford Handbook of Free Will*, ed. Robert Kane, 406-37. New York: Oxford University Press.
- Kane, Robert, 2002b, "Free Will," Oxford: Blackwell.
- Kane, Robert, 2004, "Agency, Responsibility, and Indeterminism: Reflections on Libertarian Theories of Free Will," In *Freedom and Determinism*, eds. Joseph Keim Campbell, Michael O'Rourke, and David Shier, 70-88. Cambridge, Mass.: MIT Press.
- Kane, Robert, 2005, *A Contemporary Introduction to Free Will*, New York: Oxford University Press.
- Kane, Robert. 2007a, "Libertarianism," In *Four Views on Free Will*, by John Martin Fischer, Robert Kane, Derk Pereboom, and Manuel Vargas, 5-43. Oxford: Blackwell.
- Kane, Robert. 2007b, "Response to Fischer, Pereboom, and Vargas," In *Four Views on Free Will*, by John Martin Fischer, Robert Kane, Derk Pereboom, and Manuel Vargas, 166-83. Oxford: Blackwell.

- Kane, Robert, 2009, "Reflections on Free Will, Determinism, and Indeterminism," *The Determinism and Freedom Philosophy Website*, first accessed 08/06/2009
<http://www.ucl.ac.uk/~uctytho/dfwVariousKane.html>
- Kant, Immanuel, 1784/1998, *Groundwork of the Metaphysics of Morals*, editor and translator M. Gregor, Cambridge: Cambridge University Press.
- Malm, H.M., 1989, "Killing, Letting Die, and Simple Conflicts", *Philosophy and Public Affairs*, Vol. 18, No. 3: 238-258.
- McCall, Storrs, 1994, *A Model of the Universe*, Oxford: Clarendon Press.
- McCann, Hugh J., 1998, *The Works of Agency: On Human Action, Will, and Freedom*, Ithaca: Cornell University Press.
- McKenna, Michael, 2004, "Compatibilism," *Stanford Encyclopedia of Philosophy*, Apr 26, 2004
<http://plato.stanford.edu/entries/compatibilism>
- Mele, Alfred R., 1995, *Autonomous Agents: From Self-Control to Autonomy*. New York: Oxford University Press.
- Mele, Alfred R., 1996, "Soft Libertarianism and Frankfurt-Style Scenarios." *Philosophical Topics* 24, no. 2: 123-41.
- Mele, Alfred R.; Robb, David, 1998, "Rescuing Frankfurt-Style Cases", *The Philosophical Review*, Vol. 107, No. 1.:97-112
- Mele, Alfred R., 1999a, "Kane, Luck, and the Significance of Free Will," *Philosophical Explorations*, 2: 96-104.
- Mele, Alfred R., 1999b, "Ultimate Responsibility and Dumb Luck", *Social Philosophy and Policy*, 16: 274-293.
- Mele, Alfred R., 2006, *Free Will and Luck*, Oxford University Press.
- Meyer, Susan Suave, 1993, *Aristotle on Moral Responsibility*, Cambridge, MA: Blackwell Pub.
- Mill, John Stuart (Robson, J.M. ed.), 1872 (1979), "An Examination of Sir William Hamilton's Philosophy," University of Toronto Press, Toronto.
- Moalem, Sharon and Prince, Jonathan, 2007, *Survival of the Sickest: A Medical Maverick Discovers Why We Need Disease*,
- Morris, Herbert, 1976, *On Guilt and Innocence*. Berkeley: California UP.
- Nagel, Thomas, 1976, "Moral Luck", *Proceedings of the Aristotelian Society*, Supplementary Volumes, Vol. 50: 137-151.
- Nelkin, Dana K., 2004, "Irrelevant Alternatives and Frankfurt Counterfactuals," *Philosophical Studies*, 121: 1-25.
- Nelkin, Dana K., 2008, "Moral Luck," *Stanford Encyclopedia of Philosophy*, June 3, 2008
<http://www.seop.leeds.ac.uk/entries/moral-luck/>
- Nichols, Shaun and Knobe, Joshua, 2007, "Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions," *Nous* 41, 663-685.
- Nozick, Robert, 1974, *Anarchy, State, and Utopia*, New York: Basic Books, Inc., Publishers.
- Nozick, Robert, 1981, *Philosophical Explanations*. Cambridge, Mass.: Belknap Press.
- Nussbaum, Martha, 1986, The Fragility of Goodness: Luck and Ethics in Greek Tragedy and Philosophy, Cambridge: Cambridge University Press 322-340.
- O'Connor, Timothy, 1995, "Agent Causation." *Agents, Causes, and Events: Essays on Indeterminism and Free Will*, ed. Timothy O'Connor, 173-200. New York: Oxford University Press.
- O'Connor, Timothy, 1996, "Why Agent Causation?" *Philosophical Topics* 24, no. 2: 143-58.

- O'Connor, Timothy, 1998/2001, "The Agent as Cause," in *Metaphysics: The Big Questions*, ed. P. van Inwagen and D. Zimmerman (Blackwell: Cambridge, MA).
(Reprinted in Robert Kane, ed., *Free Will*, Cambridge University Press, 2001.)
- O'Connor, Timothy, 2000a, "Causality, Mind, and Free Will," *Philosophical Perspectives* 14: 105-17
- O'Connor, Timothy, 2000b, *Persons and Causes: The Metaphysics of Free Will*. New York: Oxford University Press.
- O'Connor, Timothy, 2002, "Libertarian Views: Dualist and Agent-Causal Theories," *The Oxford Handbook of Free Will*, ed. Robert Kane, 337-55. New York: Oxford University Press.
- O'Connor, Timothy, 2005, "Freedom with a Human Face," *Midwest Studies in Philosophy*, 29: 207-27.
- Otsuka, Michael, 1998, "Incompatibilism and the Avoidability of Blame", *Ethics*, Vol. 108, No. 4: 685-701
- Pereboom, Derk, 1985, "Determinism al Dente," *Noûs*, Vol. 29, No. 1 (Mar., 1995), pp. 21-45
- Pereboom, Derk, 2000, "Alternative Possibilities and Causal Histories," *Philosophical Perspectives* 14: 119-37.
- Pereboom, Derk, 2001, *Living Without Free Will*, Cambridge: Cambridge University Press.
- Pereboom, Derk, 2002, "Living Without Free Will: The case for hard incompatibilism," *The Oxford Handbook of Free Will*, Robert Kane, ed., New York: Oxford University Press, pp. 477-88.
- Pereboom, Derk, 2005, "Defending Hard Incompatibilism," *Midwest Studies* 29, pp. 228-47.
- Pereboom, Derk. 2007. "Hard Incompatibilism." In *Four Views on Free Will*, by John Martin Fischer, Robert Kane, Derk Pereboom, and Manuel Vargas, 85-125. Oxford: Blackwell.
- Pereboom, Derk, forthcoming, "Defending Hard Incompatibilism Again," forthcoming in *Essays on Free Will and Moral Responsibility*, Nick Trakakis and Daniel Cohen, eds., Newcastle Cambridge Scholars Press.
- Pereboom, Derk, forthcoming 2, "Is Our Conception of Agent-causation Coherent?," forthcoming in *Philosophical Topics*
- Pink, Thomas, 2004, *Free Will: A Very Short Introduction*, Oxford: Oxford University Press.
- Plantinga, Alvin, 1974, *The Nature of Necessity*, New York: Oxford.
- Pojman, Louis P., 2002, "Free Will, Determinism, and Moral Responsibility: A Response to Galen Strawson", *Ethical Theory Classic and Contemporary Readings*, Wadsworth Group: 722-729.
- Popper, Karl, and Eccles, John, 1977, *The Self and Its Brain*, New York: Springer-Verlag.
- Rawls, John, 1971, *A Theory of Justice*. Cambridge, MA: Harvard University Press.
- Rachels, James, 1975, "Active and Passive Euthanasia," *The New England Journal of Medicine*, Vol. 292, pp. 78-80.
- Rachels, James, 1979, "Killing and Starving to Death," *Philosophy*, Vol. 54, No. 208, pp. 159-171.
- Rachels, James, 2003, *The Elements of Moral Philosophy Fourth Edition*, McGraw-Hill
- Reid, Thomas, 1788/1969, *Essays on the Active Powers of the Human Mind*. Cambridge, Mass.: MIT Press.
- Rescher, N., 1993, "Moral Luck", in *Moral Luck*, D. Statman (ed.), Albany: State University of New York Press.
- Richards, Norvin, 1986, "Luck and Desert", *Mind*, New Series, Vol. 95, No. 378: 198-209.

- Roberts, Jean, 1984, "Aristotle on Responsibility for Action and Character," *Ancient Philosophy*, 9: 23-36.
- Rosebury, Brian, 1995, "Moral Responsibility and Moral Luck", *Philosophical Review*, 104: 499-524.
- Roush, S., 1998, "Alternate Possibilities and Their Entertainment," *Philosophy*, Vol. 73, No. 286, pp. 559-571.
- Rowe, William L., 1991, *Thomas Reid on Freedom and Morality*. Ithaca: Cornell University Press.
- Rowe, William L., 2000, "The Metaphysics of Freedom: Reid's Theory of Agent Causation." *American Catholic Philosophical Quarterly* 74: 425-46.
- Rowe, William L., 2003, "Alternate Possibilities and Reid's Theory of Agent-Causation," In *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities*, eds. David Widerker and Michael McKenna, 219-34. Aldershot, UK: Ashgate.
- Rowe, William L., 2006, "Free Will, Moral Responsibility, and the Problem of 'Oomph'," *Journal of Ethics*, 10: 295-313.
- Scheffler, S., 2003, "What is Egalitarianism", *Philosophy and Public Affairs*, 31: 5-39.
- Searle, John R., 1980, *Rationality in Action*. Cambridge, Mass.: MIT Press. 2001.
- Searle, John R., 2004, *Freedom and Neurobiology Reflections of Free Will, Language, and Political Power*, Columbia University Press.
- Smart, J. J. C., 1961, "Free-will, Praise, and Blame," *Mind*, vol. 70, pp. 291-306.
- Smart, J. J. C., 1968, "Between Science and Philosophy," Random House, New York.
- Smart, J. J. C. and Williams, Bernard, 1973, *Utilitarianism For and Against*, Cambridge: Cambridge University Press, 1973.
- Smith, Donald and Coffman, E.J., forthcoming, "The Fall of the Mind Argument and Some Lessons about Freedom", *Topics in Contemporary Philosophy: Volume 7: Action, Ethics, and Responsibility*, MIT Press.
- Sorabji, Richard, 1980, *Necessity, Cause, and Blame: Perspectives on Aristotle's Theory*. Ithaca: Cornell University Press.
- Statman, D., 1991, "Moral and Epistemic Luck", *Ratio*, 4:146-156.
- Statman, D. 1993a, *Introduction to Moral Luck*, Albany: State University of New York Press.
- Statman, D. (ed.), 1993b, *Moral Luck*, Albany: State University of New York Press.
- Statman, D., 2005, "Doors, Keys, and Moral Luck: A Reply to Domskey," *The Journal of Philosophy*, 102: 422-436
- Strawson, Galen, 1986, *Freedom and Belief*, New York: Oxford University Press.
- Strawson, Galen, 1994/2002, "The Impossibility of Moral Responsibility." *Philosophical Studies* 75: 5-24. reprinted in *Ethical Theory Classic and Contemporary Readings* by Louis P. Pojman
- Strawson, Peter F., 1962, "Freedom and Resentment." *Proceedings of the British Academy*, XLVIII: 187-211.
- Sverdlik, S., 1988, "Crime and Moral Luck", *American Philosophical Quarterly*, 25: 79-86
- Taylor, Richard, 1966, *Action and Purpose*. Englewood Cliffs: Prentice-Hall.
- Taylor, Richard, 1992, *Metaphysics*, 4th edition. Englewood Cliffs: Prentice-Hall.
- Thomson, J. J., 1993, "Morality and Bad Luck", in *Moral Luck*, D. Statman (ed.), Albany: State University of New York Press.

- Thorp, John., 1980, *Free Will: A Defence Against Neurophysiological Determinism*. London: Routledge & Kegan Paul.
- Van Inwagen, Peter, 1983, *An Essay on Free Will*, Oxford: Clarendon Press.
- Van Inwagen, Peter, 1997, "Against Middle Knowledge," *Midwest Studies in Philosophy*, xxi: 225-36.
- Van Inwagen, Peter, 2000, "Free Will Remains a Mystery," *Philosophical Perspectives*, Vol. 14, *Action and Freedom*, 1-19.
- Van Inwagen, Peter, 2008, "How to Think about the Problem of Free Will," *The Journal of Ethics*, Vol. 12, No 3-4, 327-341
- Vihvelin, Kadri, 2007, "Arguments for Incompatibilism," *Stanford Encyclopedia of Philosophy*, Nov 13, 2007
<http://plato.stanford.edu/entries/incompatibilism-arguments/>
- Wallace, R. Jay, 1994, *Responsibility and the Moral Sentiments*. Cambridge: Harvard.
- Walker, Margaret Urban, 1991, "Moral Luck and the Virtues of Impure Agency", *Metaphilosophy*, 22: 14-27
- Walton, Kendall L., 1978, "How Remote Are Fictional Worlds from the Real World?," *The Journal of Aesthetics and Art Criticism*, Vol. 37, No. 1: 11-23
- Watson, Gary, 1975, "Free Agency," *The Journal of Philosophy*, Vol. 72, No. 8, 205-220.
- Watson, Gary, 1982, *Free Will*, New York: Oxford University Press.
- Widerker, David, 1995, "Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities," *The Philosophical Review* 104 , pp. 247-261
- Wiggins, David, 1973, "Towards a Reasonable Libertarianism." *Essays on Freedom of Action*, ed. Ted Honderich, 33-61. London: Routledge & Kegan Paul.
- Williams, Bernard, 1981, *Moral Luck*, Cambridge: Cambridge University Press.
- Williams, Bernard, 1993, "Postscript", in *Moral Luck*, D. Statman (ed.), Albany: State University of New York Press.
- Wolf, S., 1990, *Freedom Within Reason*, Oxford: Oxford University Press.
- Wolf, S., 2001, "The Moral of Moral Luck", *Philosophic Exchange*, 31: 4-19.
- Young, Robert, 1974, "Compatibilism and Freedom" *Mind*, New Series, Vol. 83, No. 32: 19-42.
- Zimmerman, Michael J., 1984, *An Essay on Human Action*. New York: Peter Lang.
- Zimmerman, Michael J., 1987, "Luck and Moral Responsibility", *Ethics*, Vol. 97, No. 2: 274-386.
- Zimmerman, Michael J., 1991, "An Essay on Moral Responsibility." *Ethics*, Vol. 101, No. 2: 408-409.
- Zimmerman, Michael J., 2002, "Taking Luck Seriously", *The Journal of Philosophy*, Vol. 99, No. 11: 553-576.
- Zimmerman, Michael J., 2006, "Moral Luck: A Partial Roadmap", *Canadian Journal of Philosophy*, Vol. 36: 585-608.